INTERNATIONAL JOURNAL

of

COMPUTERS, COMMUNICATIONS & CONTROL

With Emphasis on the Integration of Three Technologies

IJCCC
A Quarterly Journal

Year: 2008 Volume: III
Suppl. issue: Proceedings of ICCCC 2008

Agora University Editing House

CCC Publications

Licensed partner: EBSCO Publishing

www.journal.univagora.ro

Publisher & Editorial Office
CCC Publications, Agora University
Piaţa Tineretului 8, Oradea, jud. Bihor, Romania, Zip Code 410526
Tel: +40 259 427 398, Fax: +40 259 434 925, E-mail: ccc@univagora.ro
Website: www.journal.univagora.ro
ISSN 1841-9836, E-ISSN 1841-9844

International Journal of Computers, Communications and Control (IJCCC) is published from 2006 and has 4 issues/year (March, June, September, December), print & online.

Founders of IJCCC: I. Dziţac, F.G. Filip and M.J. Manolescu (2006)

This publication is subsidized by:

1. Agora University
2. The Romanian Ministry of Education and Research / The National Authority for Scientific Research

CCC Publications, powered by Agora University Publishing House, currently publishes the "International Journal of Computers, Communications & Control" and its scope is to publish scientific literature (journals, books, monographs and conference proceedings) in the field of Computers, Communications and Control.

IJCCC is indexed/abstracted/covered in a number of databases and services including:

1. ISI Thomson Scientific
2. ISJ - Journal Popularity
3. Computer & Applied Science Complete (EBSCO)
4. Current Abstracts (EBSCO Publishing)
5. Information Systems Journals (ISJ)
6. The Collection of Computer Science Bibliographies(CCSB)
7. Open J-Gate
8. FIZ KARLSRUHE's informatics portal io-port.net
9. Urlich's Periodicals Directory
10. MathSciNet
11. DOAJ
12. SCIRUS
13. Google Scholar
14. Genamics JournalSeek

Ioan Dziţac Florin Gheorghe Filip Mişu-Jan Manolescu

Editors

PROCEEDINGS OF ICCCC 2008

INTERNATIONAL CONFERENCE on

COMPUTERS, COMMUNICATIONS and CONTROL

Băile Felix, Oradea, Romania, May 15-17, 2008

www.iccc.univagora.ro
2008

Editors of the Proceedings

IOAN DZIŢAC
Agora University
8, Piaţa Tineretului,
410526 Oradea,
idzitac@univagora.ro

FLORIN GHEORGHE FILIP
Romanian Academy, Bucharest
Romanian Academy,
125, Calea Victoriei,
010071 Bucharest-1,
ffilip@acad.ro

MIŞU-JAN MANOLESCU
Agora University,
8, Piaţa Tineretului,
410526 Oradea,
rectorat@univagora.ro

Managing Editor of the Proceedings
ADRIANA MANOLESCU, Agora University, adrianamanolescu@univagora.ro

Technical Editors of the Proceedings
HOREA OROS, University of Oradea, Romania, horos@uoradea.ro

ICCCC 2008 website design
EMMA VĂLEANU, Agora University, Oradea, Romania, evaleanu@univagora.ro

ICCCC 2008
was organized by Agora University, Oradea, Romania under the aegis of

- Information Science and Technology Section of the Romanian Academy
- Forum for Knowledge Society of the Romanian Academy
- IEEE Romania Section.

Printed
Horia Neag, Metropolis SRL, Oradea, Romania, Phone. +4 0259 472640, metropolis@rdsor.ro

Sponsors
The printing of the proceedings was sponsored by the Ministry of Education and Research, Romania - National Authority for Scientific Research

Conference URL: http://www.iccc.univagora.ro

Contact: Dr. Ioan Dziţac, Agora University, Piaţa Tineretului, 8, 410526 ORADEA, ROMANIA
Phone/Fax: +40359101032, E-mail: idzitac@univagora.ro, URL: http://dzitac.rdsor.ro

ICCCC 2008 STAFF: PROGRAM COMMITTEE & ORGANIZING COMMITTEE

HONORARY CONFERENCE CHAIR
Lotfi A. ZADEH, University of California - Berkeley, USA

HONORARY PROGRAM COMMITTEE CHAIR
Florin Gheorghe FILIP, Romanian Academy, ROMANIA

GENERAL CHAIR
Ioan DZIŢAC, Agora University, Oradea, ROMANIA

CONFERENCE CHAIRS
Mişu-Jan MANOLESCU, Agora University, Oradea, ROMANIA
Nicolae ŢĂPUŞ, Politehnica University, Bucharest, ROMANIA

INTERNATIONAL PROGRAM COMMITTEE

PROGRAM CHAIRS
Ioan DZIŢAC, Agora University, Oradea, ROMANIA
Imre J. RUDAS, Budapest Tech, HUNGARY

MEMBERS
Răzvan ANDONIE, Central Washington University, USA
Grigore ALBEANU, Spiru Haret University, Bucharest, ROMANIA
Valentina E. BĂLAŞ, Aurel Vlaicu University, Arad, ROMANIA
Boldur E. BĂRBAT, Lucian Blaga University, Sibiu, ROMANIA
Barnabas BEDE, University of Texas at El Paso, USA
Vasile BERINDE, North University of Baia Mare, ROMANIA
Alexandru BICA, University of Oradea, ROMANIA
Florian Mircea BOIAN, Babes-Bolyai University, Cluj-Napoca, ROMANIA
Ganesh Dattatray BHUTKAR, Bharati Vidyapeeth University, INDIA
Daniel BREAZ, "1 Decembrie 1918" University of Alba Iulia, ROMANIA
Camelia CHIRA, Babes-Bolyai University, Cluj-Napoca, ROMANIA
Mitică CRAUS, Technical University of Iasi, ROMANIA
Paul CRISTEA, Politehnica University of Bucharest, ROMANIA
Doina DANAIATA, West University of Timisoara, ROMANIA
Ioan DESPI, University of New England, AUSTRALIA
Petre DINI, Cisco System, USA
Antonio DI NOLA, University of Salerno, ITALY
Ioan DUMITRACHE, Politehnica University of Bucharest, ROMANIA
Dan DUMITRESCU, Babes-Bolyai University, Cluj-Napoca, ROMANIA
Ömer EGECIOGLU, University of California, USA
Janos FODOR, Budapest Tech, HUNGARY
Constantin GAINDRIC, Academy of Sciences of Moldova, Republic of MOLDOVA
Xiao-Shan GAO, Academia Sinica, Beijing, CHINA
Angel GARRIDO, Facultad de Ciencias, UNED, SPAIN
Adelina GEORGESCU, University of Pitesti, ROMANIA
Bogdan GHILIC-MICU, Bucharest University of Economics, ROMANIA
Dan GRIGORAŞ, University College Cork, IRELAND
Kaoru HIROTA, Tokyo Institute of Technology, JAPAN
Afrodita IORGULESCU, Bucharest University of Economics, ROMANIA
Solomon MARCUS, IMAR, Romanian Academy, ROMANIA

ORGANIZING COMMITTEE

CHAIRS

SECRETARIAT

MEMBERS

Resume. International Conference on Computers, Communications and Control, ICCCC 2008, Băile Felix, May 15-7, 2008, is organized by Agora University of Oradea, under the aegis of Romanian Academy (Information Science and Technology Section and Forum for Knowledge Society) and IEEE Romania Section.
Satellite event: Exploratory Workshop on NL-Computation 2008 "From Natural Language to Soft Computing: New Paradigms in Artificial Intelligence" (organizer: Dan Tufiş, Research Institute for Artificial Intelligence of the Romanian Academy).
Organizing key-persons: Ioan Dziţac, Florin Gheorghe Filip, Mişu-Jan Manolescu, Adriana Manolescu, Horea Oros, Emma Văleanu.
ICCCC 2008 provides a forum for international scientists in academia and industry to present and discuss their latest research findings on a broad array of topics in computer networking and control. The Program Committee is soliciting paper describing original, previously unpublished, completed research, not currently under review by another conference or journal, addressing state-of-the-art research and development in all areas related to computer networking and control. The scope of the conference covered the following topics: Artificial Intelligence, Automata and Formal Languages, Computational Mathematics, Cryptography and Security, Control, Economic Informatics, E-Activities, Fuzzy Systems, Information Society - Knowledge Society, Natural Computing, Network Design & Internet Services, Multimedia & Communications, P2P Systems and Internet applications and Parallel and Distributed Computing.
Received manuscripts 202 (rate of acceptance and publications 54.5%).
The Program Committee received 202 submissions, originating from Algeria, Australia, Bulgaria, Chile, Egypt, France, Germany, Greece, Hungary, Italy, Japan, India, Ireland, Iraq, Iran, Macedonia, Maroc, Netherlands, Spain, Serbia, Romania, Tunisia, Turkey and USA.

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008

Contents

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 16-25

ICCCC 2008 & EWNLC 2008 Celebrates Bardeen's Centenary and Welcomes Professor Zadeh

Ioan Dziţac

Agora University, General Chair of ICCCC 2008

A substitute of preface

John Bardeen (1908-1991)
Nobel Prize for Transistor (1956)
Nobel Prize for BCS Theory (1972)

Lotfi A. Zadeh (b. 1921)
Creator of Fuzzy Set Theory (1965)
Creator of Fuzzy Logic Theory (1973)

Abstract: This edition of International Conference on Computers, Communications and Control, ICCCC 2008 [1], together with the satellite-event Exploratory Workshop on Natural Language Computation, EWNLC 2008 [2]: "From Natural Language to Soft Computing: New Paradigms in Artificial Intelligence", celebrates the Centenary of John Bardeen (1908-1991) [3-24], the co-inventor of the transistor, a very important element in the development of the computers and the communications.
ICCCC 2008 and EWNLC 2008 are honored to have a special guest as keynote speaker in the person of a famous scientist, Dr. Lotfi A. Zadeh [25-32], professor at Berkeley University of California. His Fuzzy Set Theory (1965), Fuzzy Logic Theory (1973) and the next contributions on Soft Computing (1990), Human-Machine Perception (2000) and Natural Language Computation are of a capital importance in the actual mathematics, computer science and technological applications (from the home intelligent e-devices to guiding-computers for missiles).

Other thirteen international scientists are present at this event as plenary ICCCC 2008 keynote speakers and as invited EWNLC 2008 speakers: Vasile Baltac (National School of Political Studies and Public Administration, Bucharest, Romania), Boldur Bărbat (Lucian Blaga University, Sibiu, Romania), Pierre Borne (Ecole Centrale de Lille, France), Ioan Buciu (University of Oradea, Romania), Florin Gheorghe Filip (Romanian Academy, Bucharest Romania), Janos Fodor (Budapest Tech, Hungary), Gaston Lefranc (Pontifical Catholic University of Valparaiso, Chile), Stephan Olariu (Old Dominion University, United States of America), Gheorghe Păun (Institute of Mathematics of Romanian Academy, Bucharest, Romania and University of Seville, Spain), Dragan Radojevic (Mihailo Pupin Institute, Beograd, Serbia), Athanasios D. Styliadis (ATEI, Thessaloniki, Greece), Horia-Nicolai Teodorescu (Gheorghe Asachi Technical University of Iaşi, Romania), Dan Tufiş (Research Institute for Artificial Intelligence of the Romanian Academy, Romania).
Other seven scientists will present invited lectures on the parallel sessions of these events: Marius Balas (Aurel Vlaicu University, Arad, Romania), Valentina Balas (Aurel Vlaicu University, Arad, Romania), Marius Guran (Politehnica University of Bucharest, Romania), Ştefan Iancu (Romanian Academy, Bucharest, Romania, Ioana Moisil (Lucian Blaga University, Sibiu, Romania), Grigor Moldovan (Babeş-Bolyai University, Cluj-Napoca, Romania), Gheorghe Ştefănescu (University of Illinois at Urbana-Champaign, United States of America).
During this event more than 100 papers will be presented by scientist from 21 countries: Algeria, Australia, Bulgaria, Chile, Egypt, France, Germany, Greece, Hungary, Japan, India, Ireland, Iran, Macedonia, Netherlands, Spain, Serbia, Romania, Tunisia, Turkey and United States. The papers presented at these two scientific events will be published in:

- L. A. Zadeh, D. Tufiş, F.G. Filip, I. Dziţac (eds), From Natural Language to Soft Computing: New Paradigms in Artificial Intelligence, Editing House of Romanian Academy, 2008;
- I. Dziţac, F.G. Filip, M.-J. Manolescu (eds), Proceedings of ICCCC 2008, in IJCCC, Vol.III(2008), suppl. issue, 2008.

1 John Bardeen's Bio-Sketch

"It's probably historically as important as the steam engine in the sense that the steam engine made the industrial age possible and the transistor made the information age possible" (Larry Smarr)[1].

This edition of International Conference on Computers, Communications and Control, ICCC 2008, celebrates the Centenary of John Bardeen (1908-1991), an American scientist, co-inventor of transistor and superconductivity theory.

John Bardeen (b. May 23, 1908 - d. January 30, 1991) was an American scientist (physicist and electrical engineer), who won the Nobel Prize in Physics twice: first in 1956 for the invention of the transistor (with William Shockley and Walter Brattain); and again in 1972 for a fundamental theory of conventional superconductivity (with Leon Neil Cooper and John Robert Schrieffer). Nobel Prize for Transistor (1956): The transistor revolutionized the electronics industry, allowing the Information Age to occur, and made possible the development of almost every modern e-device, from telephones and computers to missiles. In 1956, John Bardeen shared the Nobel Prize in Physics with William Shockley of Semiconductor Laboratory of Beckman Instruments and Walter Brattain of Bell Telephone Laboratories.

Nobel Prize for BCS Theory (1972): BCS theory (named after its creators, Bardeen, Cooper, and Schrieffer) explains conventional superconductivity, the ability of certain metals at low temperatures to conduct electricity without electrical resistance. BCS theory views superconductivity as a macroscopic quantum mechanical effect. In 1957, John Bardeen, in collaboration with Leon Cooper and his doctoral student John Robert Schrieffer, proposed the standard theory of superconductivity known as the BCS theory (named for their initials). Independently and at the same time, superconductivity phenomenon was explained by Nikolay Bogoliubov by means of the so-called Bogoliubov transformations. In 1972, John Bardeen shared the Nobel Prize in Physics with Leon Neil Cooper of Brown University and John Robert Schrieffer of the University of Pennsylvania for their jointly developed theory of superconductivity, usually called the BCS-theory. [3-24]

[1] http://www.jacobsschool.ucsd.edu/ lsmarr/

Other Prizes, Distinctions and Awards: In 1971, Bardeen received the IEEE Medal of Honor for "his profound contributions to the understanding of the conductivity of solids, to the invention of the transistor, and to the microscopic theory of superconductivity". Bardeen was one of 11 recipients given the Third Century Award from President George H.W. Bush in 1990 for "exceptional contributions to American society" and was granted a gold medal from the Soviet Academy of Sciences in 1988.

In 1990, John Bardeen appeared on LIFE Magazine's list of "100 Most Influential Americans of the Century."

Some information from the official webpage of the Department of Physics at the University of Illinois at Urbana-Champaign[2] :

"John Bardeen was born in Madison, Wisconsin on May 23, 1908. His father, Charles Russell Bardeen, was the first graduate of the Johns Hopkins Medical School and founder of the Medical School at the University of Wisconsin. His mother, Althea Harmer, studied oriental art at the Pratt Institute and practiced interior design in Chicago. He was one of five children.

John received his elementary and secondary education in Madison. He studied Electrical Engineering at the University of Wisconsin, receiving a B.S. in 1928 and an M.S. in 1929. The three years 1930-33 were spent doing research in geophysics at the Gulf Research Laboratories in Pittsburgh, Pennsylvania. In 1933, he returned to graduate studies in mathematical physics at Princeton University, where he had his first introduction to solid state theory from Professor E.P. Wigner, and received his Ph.D. in 1936. The three years, 1935-38, were spent as a Junior Fellow of the Society of Fellows of Harvard University, where he worked with Professors J.H. Van Vleck and P.W. Bridgeman. From 1938-41, he was an Assistant Professor at the University of Minnesota and from 1941-45, at the Naval Ordnance Laboratory in Washington, D.C. In the fall of 1945, he joined the newly formed research group in solid state physics at the Bell Telephone Laboratories, Murray Hill, New Jersey. It was there that he became interested in semiconductors and with W.H. Brattain discovered the transistor effect in late 1947. He left Bell Labs in 1951 to become Professor of Electrical Engineering and of Physics at the University of Illinois, Urbana, where he was Professor and Emeritus Professor.

At Illinois, Bardeen established two major research programs, one in the Electrical Engineering Department dealing with both experimental and theoretical aspects of semiconductors, and one in the Physics Department which dealt with theoretical aspects of macroscopic quantum systems, particularly superconductivity and quantum liquids. The microscopic theory of superconductivity, developed in collaboration with L.N. Cooper and J.R. Schrieffer in 1956 and 1957, has had profound implications for nearly every field of physics from elementary particle to nuclear and the helium liquids to neutron stars. During his sixty year scientific career, he made significant contributions to almost every aspect of condensed matter physics from his early work on the electronic behavior of metals, the surface properties of semiconductors and the theory of diffusion of atoms in crystals to his most recent work on quasi-one-dimensional metals. In his eighty-third year, he continued to publish original scientific papers.

During this period, Bardeen maintained active interests in engineering and technology. He began consulting for Xerox Corporation in 1951, when it was still called Haloid and the Research Department was located in a frame house in Rochester, New York. He worked with Xerox throughout their spectacular development, and later served on the Xerox Board of Directors. He also consulted with General Electric Corporation for many years and with several other technology firms.

Bardeen, a Fellow of the American Physical Society, served on the Council from 1954-57 and was President in 1968-69. He was elected to the National Academy of Sciences in 1954 and the National Academy of Engineering in 1972. He served on the U.S. President's Science Advisory Committee from 1959 to 1962 and on the White House Science Council in 1981-82. He was a founding member of the Commission on Very Low Temperatures of the International Union of Pure and Applied Physics from 1963-1972, serving as chairman in 1969-1972. From 1961-1974 he was a member of the Board of Directors of Xerox Corporation and was a member of the Board of Supertex, Inc. from 1983 to 1991." [6][3]

2 Lotfi A. Zadeh's Bio-Sketch

Lotfi A. Zadeh joined the Department of Electrical Engineering at the University of California, Berkeley, in 1959, and served as its chairman from 1963 to 1968. Earlier, he was a member of the electrical engineering faculty at Columbia University. In 1956, he was a visiting member of the Institute for Advanced Study in Princeton, New Jersey. In addition, he held a number of other visiting appointments, among them a visiting professorship in

[2]http://www.physics.uiuc.edu/history/bardeen.htm
[3]http://www.physics.uiuc.edu/history/bardeen.htm

Electrical Engineering at MIT in 1962 and 1968; a visiting scientist appointment at IBM Research Laboratory, San Jose, CA, in 1968, 1973, and 1977; and visiting scholar appointments at the AI Center, SRI International, in 1981, and at the Center for the Study of Language and Information, Stanford University, in 1987-1988. Currently he is a Professor in the Graduate School, and is serving as the Director of BISC (Berkeley Initiative in Soft Computing).

Until 1965, Dr. Zadeh's work had been centered on system theory and decision analysis. Since then, his research interests have shifted to the theory of fuzzy sets and its applications to artificial intelligence, linguistics, logic, decision analysis, control theory, expert systems and neural networks. Currently, his research is focused on fuzzy logic, soft computing, computing with words, and the newly developed computational theory of perceptions and precisiated natural language.

An alumnus of the University of Tehran, MIT, and Columbia University, Dr. Zadeh is a fellow of the IEEE, AAAS, ACM, AAAI and IFSA, and a member of the National Academy of Engineering. He held NSF Senior Postdoctoral Fellowships in 1956-57 and 1962-63, and was a Guggenheim Foundation Fellow in 1968. Dr. Zadeh was the recipient of the IEEE Education Medal in 1973 and a recipient of the IEEE Centennial Medal in 1984. In 1989, Dr. Zadeh was awarded the Honda Prize by the Honda Foundation, and in 1991 received the Berkeley Citation, University of California.

In 1992, Dr. Zadeh was awarded the IEEE Richard W. Hamming Medal "For seminal contributions to information science and systems, including the conceptualization of fuzzy sets." He became a Foreign Member of the Russian Academy of Natural Sciences (Computer Sciences and Cybernetics Section) in 1992, and received the Certificate of Commendation for AI Special Contributions Award from the International Foundation for Artificial Intelligence. Also in 1992, he was awarded the Kampe de Feriet Prize and became an Honorary Member of the Austrian Society of Cybernetic Studies.

In 1993, Dr. Zadeh received the Rufus Oldenburger Medal from the American Society of Mechanical Engineers "For seminal contributions in system theory, decision analysis, and theory of fuzzy sets and its applications to AI, linguistics, logic, expert systems and neural networks." He was also awarded the Grigore Moisil Prize for Fundamental Researches, and the Premier Best Paper Award by the Second International Conference on Fuzzy Theory and Technology. In 1995, Dr. Zadeh was awarded the IEEE Medal of Honor "For pioneering development of fuzzy logic and its many diverse applications." In 1996, Dr. Zadeh was awarded the Okawa Prize "For outstanding contribution to information science through the development of fuzzy logic and its applications."

In 1997, Dr. Zadeh was awarded the B. Bolzano Medal by the Academy of Sciences of the Czech Republic "For outstanding achievements in fuzzy mathematics." He also received the J.P. Wohl Career Achievement Award of the IEEE Systems, Science and Cybernetics Society. He served as a Lee Kuan Yew Distinguished Visitor, lecturing at the National University of Singapore and the Nanyang Technological University in Singapore, and as the Gulbenkian Foundation Visiting Professor at the New University of Lisbon in Portugal. In 1998, Dr. Zadeh was awarded the Edward Feigenbaum Medal by the International Society for Intelligent Systems, and the Richard E. Bellman Control Heritage Award by the American Council on Automatic Control. In addition, he received the Information Science Award from the Association for Intelligent Machinery and the SOFT Scientific Contribution Memorial Award from the Society for Fuzzy Theory in Japan. In 1999, he was elected to membership in Berkeley Fellows and received the Certificate of Merit from IFSA (International Fuzzy Systems Association). In 2000, he received the IEEE Millennium Medal; the IEEE Pioneer Award in Fuzzy Systems; the ASPIH 2000 Lifetime Distinguished Achievement Award; and the ACIDCA 2000 Award for the paper, "From Computing with Numbers to Computing with Words-From Manipulation of Measurements to Manipulation of Perceptions." In addition, he received the Chaos Award from the Center of Hyperincursion and Anticipation in Ordered Systems for his outstanding scientific work on foundations of fuzzy logic, soft computing, computing with words and the computational theory of perceptions. In 2001, Dr. Zadeh received the ACM 2000 Allen Newell Award for seminal contributions to AI through his development of fuzzy logic. In addition, he received a Special Award from the Committee for Automation and Robotics of the Polish Academy of Sciences for his significant contributions to systems and information science, development of fuzzy sets theory, fuzzy logic control, possibility theory, soft computing, computing with words and computational theory of perceptions. In 2003, Dr. Zadeh was elected as a foreign member of the Finnish Academy of Sciences, and received the Norbert Wiener Award of the IEEE Society of Systems, Man and Cybernetics "For pioneering contributions to the development of system theory, fuzzy logic and soft computing." In 2004, Dr. Zadeh was awarded Civitate Honoris Causa by Budapest Tech (BT) Polytechnical Institution, Budapest, Hungary. Also in 2004, he was awarded the V. Kaufmann Prize by the International Association for Fuzzy-Set Management and Economy (SIGEF). In 2005, Dr. Zadeh was elected as a foreign member of Polish Academy of Sciences, Korea Academy of Science & Technology and Bulgarian Academy of Sciences. He was also awarded the Nicolaus Copernicus Medal of the Polish Academy of Sciences and the J. Keith

Brimacombe IPMM Award.

Dr. Zadeh is a recipient of twenty-three honorary doctorates from: Paul-Sabatier University, Toulouse, France; State University of New York, Binghamton, NY; University of Dortmund, Dortmund, Germany; University of Oviedo, Oviedo, Spain; University of Granada, Granada, Spain; Lakehead University, Canada; University of Louisville, KY; Baku State University, Azerbaijan; the Silesian Technical University, Gliwice, Poland; the University of Toronto, Toronto, Canada; the University of Ostrava, the Czech Republic; the University of Central Florida, Orlando, FL; the University of Hamburg, Hamburg, Germany; the University of Paris(6), Paris, France; Johannes Kepler University, Linz, Austria; University of Waterloo, Canada; the University of Aurel Vlaicu, Arad, Romania; Lappeenranta University of Technology, Lappeenranta, Finland; Muroran Institute of Technology, Muroran, Japan; Hong Kong Baptist University, Hong Kong, China; Indian Statistical Institute, Kolkata, India; University of Saskatchewan, Saskatoon, Canada and the Polytechnic University of Madrid, Madrid, Spain.

Dr. Zadeh has single-authored over two hundred papers and serves on the editorial boards of over fifty journals. He is a member of the Advisory Committee, Center for Education and Research in Fuzzy Systems and Artificial Intelligence, Iaşi, Romania; Senior Advisory Board, International Institute for General Systems Studies; the Board of Governors, International Neural Networks Society; and is the Honorary President of the Biomedical Fuzzy Systems Association of Japan and the Spanish Association for Fuzzy Logic and Technologies. In addition, he is a member of the Advisory Board of the National Institute of Informatics, Tokyo; a member of the Governing Board, Knowledge Systems Institute, Skokie, IL; and an honorary member of the Academic Council of NAISO-IAAC.

Address: Lotfi A. Zadeh
Professor in the Graduate School and Director,
Berkeley Initiative in Soft Computing (BISC),
Computer Science Division, Department of EECS,
University of California, Berkeley, CA 94720-l776;
Telephone: 5l0-642-4959; Fax: 5l0-642-l7l2;
E-mail: zadeh@eecs.berkeley.edu; http://www.cs.berkeley.edu/ zadeh/

3 International Conference on Computers, Communications & Control

3.1 ICCCC 2006

The first edition of this conference (in new format [4]), ICCCC 2006, had been organized by Ioan Dziţac from Agora University (President Misu-Jan Manolescu) under the aegis of IEEE Romania Section (President Nicolae Tăpuş) and under the guidance of acad. Florin Gheorghe Filip, vice-president of Romanian Academy.

Resume. International Conference on Computers, Communications and Control, Băile Felix, June 1-3, 2006, organized by Agora University of Oradea, under the aegis of IEEE Romania Section. Satellite event: Launch of

[4]This conference is an extension of International Conference on Computers and Communications (ICCC 2004, founded by I. Dziţac, C. Popescu, F.G. Filip and H. Oros) in Control field, performed in 2006 by I. Dziţac, F.G. Filip and M.-J. Manolescu. Edition ICCC 2004: International Conference on Computers and Communications, Băile Felix, May 27-29, 2004, organized by University of Oradea, under the aegis of Romanian Society of Applied and Industrial Mathematics (ROMAI). Satellite event: Open Workshop on European ICT Qualifications (organizer: Willi Petersen, University of Flensburg, Germany). Organizing key-persons: Ioan Dziţac, Florin Gheorghe Filip, Horea Oros, Constantin Popescu, Eugen Petac, Willi Petersen. Received manuscripts 112 (rate of acceptance and publications 59.8%). Geographical area of participants: Australia, Austria, China, Egypt, Finland, France, Germany, Greece, India, Ireland, Italy, Japan, Moldova, Romania, Spain, United Kingdom, United States.

Plenary invited speakers: Antonio Di Nola (University of Salerno, Italy), Gheorghe Păun (IMAR Romania & University of Seville, Spain), Horia-Nicolai Teodorescu (Gheorghe Asachi Tech. University of Iaşi), A. Willi Petersen (University of Flensburg, Germany).

Publications:

- Ioan Dziţac, Teodor Maghiar, Constantin Popescu (eds.), - *Proceedings of International Conference on Computers and Communications (ICCC 2004)*, 27-29 May, 2004, Baile Felix- Oradea, Ed. Univ. din Oradea, ISBN 973 - 613 - 542 - X (2004), 434 p (67 papers);
- Studies in Informatics and Control - SIC, Vol.14(2005), no.1 (6 papers);
- Fuzzy Systems & A.I., Reports and Letters, Vol. 10, Nos. 1-2, 2005 (5 papers).

International Journal of Computers, Communications and Control, IJCCC (organizer: Ioan Dziţac, Agora University, Oradea, Romania).
Organizing key-persons: Ioan Dziţac, Florin Gheorghe Filip, Mişu-Jan Manolescu, Adriana Manolescu, Horea Oros.
Received manuscripts 142 (rate of acceptance and publications 64%).

Geographical area of participants: Algeria, France, Germany, Greece, Hungary, Italy, India, Ireland, Japan, Moldova, Romania, Serbia, Spain, Thailand, Tunisia, United States.

Plenary invited speakers: Paul Dan Cristea (Politehnica University of Bucharest, Romania), Gabriel Ciobanu (Gheorghe Asachi Tech. University of Iaşi), Janos Fodor (Budapest Tech, Hungary), Dan Tufiş (Research Institute for Artificial Intelligence of the Romanian Academy, Romania).

Publications:

- Ioan Dziţac, Florin Gheorghe Filip, Mişu-Jan Manolescu (eds.), Proceedings of ICCCC 2006, in International Journal of Computers, Communications & Control (IJCCC), Vol. I (2006), supplementary issue, ISSN 1841-9836, 512 p (84 papers);
- International Journal of Computers, Communications and Control (11 papers).

3.2 ICCCC 2008

The second edition, ICCCC 2008, is organized by Ioan Dziţac from Agora University (President Misu-Jan Manolescu) under the aegis of Romanian Academy (Information Science and Technology Section- President acad. Mihai Drăgănescu and Forum for Knowledge Society - President acad. Florin Gheorghe Filip) and IEEE Romania Section (President prof. Nicolae Tapus).

ICCCC 2008 provides a forum for international scientists in academia and industry to present and discuss their latest research findings on a broad array of topics in computer networking and control. The Program Committee had solicited papers describing original, previously unpublished, completed research, not currently under review by another conference or journal, addressing state-of-the-art research and development in all areas related to computer networking and control.

The scope of the conference covered the following topics: Artificial Intelligence, Automata and Formal Languages, Computational Mathematics, Cryptography and Security, Control, Economic Informatics, E-Activities, Fuzzy Systems, Information Society - Knowledge Society, Natural Computing, Network Design & Internet Services, Multimedia & Communications, P2P Systems and Internet applications and Parallel and Distributed Computing.

The Program Committee received 202 submissions (rate of acceptance and publications 54.5%), originating from Algeria, Australia, Bulgaria, Chile, Egypt, France, Germany, Greece, Hungary, Italy, Japan, India, Ireland, Iraq, Iran, Macedonia, Maroc, Netherlands, Romania, Spain, Serbia, Tunisia, Turkey and USA.

Each submission was reviewed by two Program Committee members, or other experts. Out of the 202 papers only 110 (54%) were accepted for presentation at the conference and for publication.

Resume. International Conference on Computers, Communications and Control, Băile Felix, May 15-7, 2008, organized by Agora University of Oradea, under the aegis of Romanian Academy (Information Science and Technology Section and Forum for Knowledge Society) and IEEE Romania Section.

Satellite event: Exploratory Workshop on NL-Computation 2008 "From Natural Language to Soft Computing: New Paradigms in Artificial Intelligence" (organizer: Dan Tufiş, Research Institute for Artificial Intelligence of the Romanian Academy).

Organizing key-persons: Ioan Dziţac, Florin Gheorghe Filip, Mişu-Jan Manolescu, Adriana Manolescu, Horea Oros, Emma Văleanu.

4 Exploratory Workshop on Natural Language Computation: "From Natural Language to Soft Computing: New Paradigms in Artificial Intelligence" (EWNLC 2008)

Exploratory Workshop on Natural Language Computation (EWNLC 2008) is organized as satellite-event of ICCCC 2008 by Dr. Dan Tufiş - director of Institute of Artificial Intelligence of Romanian Academy - and Dr. Ioan Dziţac - Agora University, and is addressed to the 30 participants from Romania and other 5 countries that

were selected from over 100 participants at ICCCC 2008: scientific researchers, PhD and PhD students, from universities and research units. The selection of the participants was based on the analysis of the scientific experience and the capacity to actively participate to the debates of each of the 100 participants who submitted papers for ICCCC 2008. Some other criteria were also used: geographical distribution (Chile, France, Greece, Hungary, Romania (Arad, Bucureşti, Cluj-Napoca, Iaşi, Oradea, Sibiu, Târgovişte), Serbia, US); repartition according to the age (experienced researchers - 24 PhDs and young researchers - 6 PhD candidates) and sex (half of the participants are women, both experienced and young researchers); the repartition among universities and other research centers; repartition based on the domain of expertise of potential participants (electronics and telecommunications, computer science, artificial intelligence, economical informatics, automatics, mathematics, economy, communication sciences, robotics, medical informatics).

EWNLC 2008 target is to promote the scientific collaboration between researchers from different countries, with different professional experiences, in order to apply for international projects. Natural Language Computation (NL-Computation or NLC) is a field of research initiated and promoted by Lotfi A. Zadeh, main key-speaker of this exploratory workshop (EW).

The key idea of the workshop is excellent defined in the words of professor Zadeh in the following quote: "Computation with information described in natural language is closely related to Computing with Words. NL-Computation is of intrinsic importance because much of human knowledge is described in natural language. This is particularly true in such fields as economics, data mining, systems engineering, risk assessment and emergency management. It is safe to predict that as we move further into the age of machine intelligence and mechanized decision-making, NL-Computation will grow in visibility and importance." (Lotfi A. Zadeh, 2007)

The keywords of this workshop are Natural Language Computation (NL-Computation), Generalized Constraint Language (GCL), Granular Computing (GC), Generalized Theory of Uncertainty (GTU), Soft Computing, and Artificial Intelligence (AI).

The key lecture of EWNLC 2008 is "*A New Frontier in Computation - Computation Described in Natural Language*", by Lotfi A. Zadeh.

In the last decades the calculation possibilities have extended on the imprecise size, because of the fuzzy sets theory, of the fuzzy logic and what we call today "soft computing", concepts introduced by professor Lotfi A. Zadeh. A natural language is essentially a perception description system. Perceptions such as estimating distances, weights, heights, temperature, as well as most of the properties of physical and mental objects, are inherently vague, reflecting the limited ability of the sensory organs, and mostly of the brain, to solve in a strict categorical manner and to store the information as such. From this perspective, "the fuzziness of natural languages is a direct consequence of the fuzziness of perceptions"[5]

In the last years professor Zadeh developed a theory concerning the representation of natural language like a computing language. One fundamental concept of this theory is "precisiation" of natural language, in the sense of transforming it in a precise, formal construct. The precisiated natural language (PNL) is the result of the transformations of natural language constructions into constructions of a Generalized Constraint Language (GCL). The expressive power of GCL is far greater than that of other AI languages (LISP, PROLOG) or other conventional logic-based meaning-representation languages (predicate logic, modal logic, a.s.o.). The main reason of this research is that most of the applications based on Natural Language Processing (semantic document categorisation, automatic text summarization, human-machine dialogue, machine translation) could be redefined in terms of GCL representations, with the advantage of a more precise processing of the perceptual information and of a more direct approach to the cognitive objectives of AI.

Professor Zadeh accepted to give a lecture at the *International Conference on Computers, Communications and Control*, ICCCC 2008, organized during May 15 and 17 at Baile Felix, by the Agora University of Oradea under the auspices of the Romanian Academy. Other persons agreed to participate to this manifestation: researchers interested in this field, from Romania (PhD professors, young PhD candidates, and also 4 members of the Romanian Academy) and also from prestigious Universities and research centres in the world. Hence it naturally appeared the idea to create this opportunity: to organize an investigating workshop on "Natural Language & Soft Computing for Artificial Intelligence" that will permit to create an international research network in this domain, where Romanian researchers to be involved and cooperate.

[5]Zadeh, L. A. 1999. From Computing with Numbers to Computing with Words–From Manipulation of Measurements to Manipulation of Perceptions. IEEE Transactions on Circuits and Systems 45(1): 105-119

Objective

This workshop, that will deal with the calculation described in natural language, is an exploratory one, integrated with the activities of the ICCCC 2008 Conference, meant for participants from Romania and other countries (researchers, PhD candidates and graduates from Universities and Research Centres) from the domain of Natural Language Processing and Computation based on a precisiated natural language. The main objective of the workshop is to promote the inter-disciplinary collaboration among the researchers from different countries having different professional experiences, with the goal of applying for international research projects in the domains of theory and practice of natural languages and PNL-type computation.

State of the art

The current research and trends in the NLC domain are clearly defined in the passage below (a quote from L. Zadeh) and are sustained by the references included.

"In conventional modes of computation, the objects of computation are values of variables. In computation with information described in natural language, or NL-Computation for short, the objects of computation are not values of variables but states of information about the values of variables, with the added assumption that information is described in natural language. A simple example: X is a real-valued random variable. What I know about X is: (a) usually X is much larger than approximately a; and (b) usually X is much smaller than approximately b, with a less than b. What is the expected value of X? Another simple example: (a) overeating causes obesity; and (b) obesity causes high blood pressure. I overeat. What is the probability that I will develop high blood pressure? The importance of NL-Computation derives from the fact that much of human knowledge, and particularly world knowledge, is expressed in natural language. Natural languages are intrinsically imprecise. A prerequisite to computation is precisiation, that is, translation of the given information into what is referred to as precisiation language. A key idea in the approach which is described is that of representing information as a so-called generalized constraint. This idea serves to construct an expressive precisiation language called the Generalized Constraint Language (GCL), whose elements are generalized constraints. Propositions and predicates drawn from a natural language are precisiated via translation into GCL. Through precisiation, computation with information described in natural language is reduced to computation with generalized constraints."[6]

"It is a deep-seated tradition in science to view uncertainty as a province of probability theory. The generalized theory of uncertainty (GTU) which is outlined in this paper breaks with this tradition and views uncertainty in a much broader perspective. Uncertainty is an attribute of information. A fundamental premise of GTU is that information, whatever its form, may be represented as what is called a generalized constraint. The concept of a generalized constraint is the centerpiece of GTU. In GTU, a probabilistic constraint is viewed as a special-albeit important-instance of a generalized constraint.A generalized constraint is a constraint of the form X isr R, where X is the constrained variable, R is a constraining relation, generally non-bivalent, and r is an indexing variable which identifies the modality of the constraint, that is, its semantics. The principal constraints are: possibilistic (r=blank); probabilistic (r=p); veristic (r=v); usuality (r=u); random set (r=rs); fuzzy graph (r=fg); bimodal (r=bm); and group (r=g). Generalized constraints may be qualified, combined and propagated. The set of all generalized constraints together with rules governing qualification, combination and propagation constitutes the generalized constraint language (GCL).The generalized constraint language plays a key role in GTU by serving as a precisiation language for propositions, commands and questions expressed in a natural language. Thus, in GTU the meaning of a proposition drawn from a natural language is expressed as a generalized constraint. Furthermore, a proposition plays the role of a carrier of information. This is the basis for equating information to a generalized constraint.In GTU, reasoning under uncertainty is treated as propagation of generalized constraints, in the sense that rules of deduction are equated to rules which govern propagation of generalized constraints. A concept which plays a key role in deduction is that of a protoform (abbreviation of prototypical form). Basically, a protoform is an abstracted summary-a summary which serves to identify the deep semantic structure of the object to which it applies. A deduction rule has two parts: symbolic-expressed in terms of protoforms-and computational.GTU represents a significant change both in perspective and direction in dealing with uncertainty and information. The concepts and techniques introduced in this paper are illustrated by a number of examples."[7]

[6]Zadeh, Lotfi A., **New Frontier in Computation: Computation with Information Described in Natural Language**, Information Reuse and Integration, 2007. IRI 2007. [14]

[7]Zadeh, Lotfi A., Information Sciences-Informatics and Computer Science: An International Journal,Volume 172 , Issue 1-2 (June 2005), pp. 1 - 40, 2005

Acknowledgment

I would like to thank to all the collaborators for their support. Without their help I would have not been able to organize this edition of ICCCC 2008.

We would like to thank the members of the Program Committee (co-chair Imre Rudas), the additional reviewers and the members of the Organizing Committee (chairs Răzvan Andonie and Adriana Manolescu, secretaries Horea Oros and Emma Văleanu) for their work and support.

On behalf of the Program Committee I would like to gratefully acknowledge all authors who submitted papers for theirs effort in increasing the scientific standards of the this edition of our conference.

We express our gratitude to our sponsors for their financial support:

- CNCSIS (president Ioan Dumitrache);
- ANCS (president Anton Anton);
- BitDefender (general manager Florin Talpeş);
- CONIZ ROMARG (general manager Gheorghe Griguţa);
- SoftNet (general manager Vasile Baltac);
- ARCER (general manager Gheorghe Ofrim);
- ADETRANS (manager general Ioan Drimuş);
- Banca Transilvania-Oradea (manager general Lucia Pojoca).

References

[1] http://www.iccc.univagora.ro/

[2] http://www.iccc.univagora.ro/?page=workshop

[3] http://en.wikipedia.org/wiki/John_Bardeen

[4] http://www.jacobsschool.ucsd.edu/ lsmarr/index.html

[5] John Bardeen, Nobelist, Inventor of Transistor, Dies", Washington Post, 1991-01-31

[6] http://www.physics.uiuc.edu/history/bardeen.htm

[7] http://www.library.uiuc.edu/archives/ead/ua/1110020/1110020f.html

[8] http://nobelprize.org/nobel_prizes/physics/laureates/1972/bardeen-bio.html

[9] http://www.nobel-winners.com/Physics/john_bardeen.html

[10] http://www.ieee.org/web/aboutus/history_center/

[11] http://nobelprize.org/nobel_prizes/physics/laureates/1956/bardeen-bio.html

[12] http://www.pbs.org/transistor/album1/bardeen/

[13] http://www.thocp.net/biographies/bardeen_john.html

[14] http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4296570/4296571/04296579.pdf

[15] http://people.clarkson.edu/ ekatz/scientists/bardeen.htm

[16] http://johnbardeen.com/

[17] http://www.britannica.com/eb/article-9013337/John-Bardeen

[18] http://www.library.uiuc.edu/archives/ead/ua/1110020/1110020f.html

[19] http://inventors.about.com/library/weekly/aa061698.htm

[20] http://almaz.com/nobel/physics/1972a.html

[21] http://www.pbs.org/transistor/background1/events/miraclemo.html

[22] http://www.library.uiuc.edu/archives/ead/ua/1110020/1110020series7.html

[23] http://www.nap.edu/openbook.php?isbn=0309048478&page=2

[24] http://www.nap.edu/catalog.php?record_id=10372

[25] http://www.cs.berkeley.edu/ zadeh/

[26] http://www-bisc.cs.berkeley.edu/

[27] http://zadeh.cs.berkeley.edu/

[28] http://en.wikipedia.org/wiki/Lotfi_Asker_Zadeh

[29] http://www.azer.com/aiweb/categories/magazine/24_folder/24_articles/24_fuzzylogic.html

[30] http://www.ieee.org/web/aboutus/history_center/biography/zadeh.html

[31] http://www.eecs.berkeley.edu/Faculty/Homepages/zadeh.html

[32] http://www.virtualarad.net/news/2003/va_n050703_ro.htm

Ioan Dziţac
Economic Informatics Department
Agora University
E-mail: idzitac@univagora.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 26-27

A New Frontier in Computation - Computation with Information Described in Natural Language

Plenary invited paper & workshop invited key lecture

Lotfi A. Zadeh

Department of EECS, University of California, Berkeley, CA 94720-1776;
Telephone: 510-642-4959; Fax: 510-642-1712; E-Mail: zadeh@eecs.berkeley.edu.
Research supported in part by ONR N00014-02-1-0294, BT Grant CT1080028046, Omron Grant, Tekes Grant, Chevron Texaco Grant and the BISC Program of UC Berkeley.

Extended Abstract

What is meant by Computation with Information Described in Natural Language, or NL-Computation, for short? Does NL-Computation constitute a new frontier in computation? Do existing bivalent-logic-based approaches to natural language processing provide a basis for NL-Computation? What are the basic concepts and ideas which underlie NL-Computation? These are some of the issues which are addressed in the following.

What is computation with information described in natural language? Here are simple examples. I am planning to drive from Berkeley to Santa Barbara, with stopover for lunch in Monterey. It is about 10 am. It will probably take me about two hours to get to Monterey and about an hour to have lunch. From Monterey, it will probably take me about five hours to get to Santa Barbara. What is the probability that I will arrive in Santa Barbara before about six pm? Another simple example: A box contains about twenty balls of various sizes. Most are large. What is the number of small balls? What is the probability that a ball drawn at random is neither small nor large? Another example: A function, f, from reals to reals is described as: If X is small then Y is small; if X is medium then Y is large; if X is large then Y is small. What is the maximum of f? Another example: Usually the temperature is not very low, and usually the temperature is not very high. What is the average temperature? Another example: Usually most United Airlines flights from San Francisco leave on time. What is the probability that my flight will be delayed?

Computation with information described in natural language is closely related to Computing with Words. NL-Computation is of intrinsic importance because much of human knowledge is described in natural language. This is particularly true in such fields as economics, data mining, systems engineering, risk assessment and emergency management. It is safe to predict that as we move further into the age of machine intelligence and mechanized decision-making, NL-Computation will grow in visibility and importance.

Computation with information described in natural language cannot be dealt with through the use of machinery of natural language processing. The problem is semantic imprecision of natural languages. More specifically, a natural language is basically a system for describing perceptions. Perceptions are intrinsically imprecise, reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. Semantic imprecision of natural languages is a concomitant of imprecision of perceptions.

Our approach to NL-Computation centers on what is referred to as generalized-constraint-based computation, or GC-Computation for short. A fundamental thesis which underlies NL-Computation is that information may be interpreted as a generalized constraint. A generalized constraint is expressed as X isr R, where X is the constrained variable, R is a constraining relation and r is an indexical variable which defines the way in which R constrains X. The principal constraints are possibilistic, veristic, probabilistic, usuality, random set, fuzzy graph and group. Generalized constraints may be combined, qualified, propagated, and counter propagated, generating what is called the Generalized Constraint Language, GCL. The key underlying idea is that information conveyed by a proposition may be represented as a generalized constraint, that is, as an element of GCL.

In our approach, NL-Computation involves three modules: (a) Precisiation module; (b) Protoform module; and (c) Computation module. The meaning of an element of a natural language, NL, is precisiated through translation into GCL and is expressed as a generalized constraint. An object of precisiation, p, is referred to as precisiend, and the result of precisiation, p^*, is called a precisiand. Usually, a precisiend is a proposition, a system of propositions or a concept. A precisiend may have many precisiands. Definition is a form of precisiation. A precisiand may be viewed as a model of meaning. The degree to which the intension (attribute-based meaning) of p* approximates

to that of p is referred to as cointension. A precisiand, p^*, is cointensive if its cointension with p is high, that is, if p^* is a good model of meaning of p.

The Protoform module serves as an interface between Precisiation and Computation modules. Basically, its function is that of abstraction and summarization.

The Computation module serves to deduce an answer to a query, q. The first step is precisiation of q, with precisiated query, q^*, expressed as a function of n variables $u_1, \ldots, u_n$. The second step involves precisiation of query-relevant information, leading to a precisiand which is expressed as a generalized constraint on $u_1, \ldots, u_n$. The third step involves an application of the extension principle, which has the effect of propagating the generalized constraint on $u_1, \ldots, u_n$ to a generalized constraint on the precisiated query, q^*. Finally, the constrained q^* is interpreted as the answer to the query and is retranslated into natural language.

The generalized-constraint-based computational approach to NL-Computation opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories. Particularly important application areas are decision-making with information described in natural language, economics, systems engineering, risk assessment, qualitative systems analysis, search, question-answering and theories of evidence.

Lotfi A. Zadeh
Professor in the Graduate School and Director
Berkeley Initiative in Soft Computing (BISC), Computer Science Division
Department of EECS, University of California
Berkeley, CA 94720-l776; Telephone: 5l0-642-4959; Fax: 5l0-642-l7l2
E-mail: zadeh@eecs.berkeley.edu
http://www.cs.berkeley.edu/ zadeh/

Research supported in part by ONR N00014-02-1-0294, BT Grant CT1080028046, Omron Grant, Tekes Grant and the BISC Program of UC Berkeley.

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 28-32

World Knowledge for Control Applications by Fuzzy-Interpolative Systems

Parallel session invited paper

Marius M. Bălaş, Valentina E. Bălaş

Abstract: The paper is discussing the necessity of providing controllers with incipient elements of world knowledge: general knowledge on system theory, specific knowledge on the processes, etc. This can be done by means of the fuzzy-interpolative systems, allied with simulation models and/or planners. A structure of world knowledge embedding planned controller is illustrating the idea.
Keywords: fuzzy-interpolative controllers fuzzy-interpolative expert systems, knowledge embedding by computer models, planners.

1 Introduction

In ref. [1] Lotfi A. Zadeh affirmed that the main weakness of the Question-Answering Systems is the absence of the world knowledge. World knowledge WK is the knowledge acquired through experience, education and communication.

The components of WK are [1]:

- Propositional: Paris is the capital of France;
- Conceptual: Climate;
- Ontological: Rainfall is related to climate;
- Existential: A person cannot have more than one father;
- Contextual: Tall.

Some of the main characteristics of WK are:

- Much of WK is perception-based;
- Much of WK is negative, i.e., relates to impossibility or nonexistence;
- Much of WK is expressed in a natural language.

Obviously KW is highly necessary to the human emulating AI products. Nevertheless this approach must overcome lots of difficulties: WK need huge memory capacity, the representation techniques must be in the same time comprehensive, specific and portable, the selection of the knowledge, the learning and the forgetting processes need further fundamental conceptual investigations, etc.

The aim of this paper is to answer the following question: "Can low level computing devices: μP, μC, *DSP*, etc. benefit of WK, when even the sophisticated modern AI software, running on powerful workstations, is encountering difficulties?"

2 The Fuzzy-Interpolative Systems

A *fuzzy-interpolative controller* FIC is a fuzzy controller that can be equaled with a corresponding look-up table with linear interpolations. The FIC concept must not be confounded with the *fuzzy rule interpolation*, originally introduced by L.T. Kóczy and K. Hirota [2], [3].

A typical FIC is a Sugeno controller with triangular or trapezoidal fuzzy partitions, prod-sum inference and COG defuzzyfication [4], [5], [6], [7], etc. The interpolative counterpart of this controller is the look-up table with linear interpolations (as the corresponding Simulink-Matlab block). FICs started from the practical observation that the Matlab FIS (Fuzzy Inference System) toolkit is demanding notable resources and occasionally it encounters

computational problems, while an equivalent look-up-table performs almost instantly, although they are producing the same control surface.

The fundamental advantage of FICs is the easiness of their implementation. In high level programming languages the look-up tables bring effectiveness, resources saving and quick developments. In fact the interpolative implementations are immediate in any possible software technology (even ASM) since the interpolation networks can be directly associated to addressable memories. These way fuzzy interpolative expert systems can be implemented virtually in any software technology. Digital hardware circuits (μC, DSPs) can also implement FICs due to their memory type architecture. However the most outstanding feature of the interpolative systems is their compatibility with the analog hardware technologies. Some possible analog technologies were mentioned, such as the translinear analog CMOS [4] and even nanometric circuits [8].

In the same time, using the fuzzy theoretical perspective, sophisticated applications become feasible. This is the case of the *fuzzy self-adaptive interpolative controllers* FSAIC [4], [5].

In close loop control applications FICs are perfectly matching a fundamental time analyze tool: *the phase trajectory of the error* and their specific analyze method, the qualitative analyze. We can rely on the figure 1 succession of theoretical tools that are involved into the conception, the development and the implementation of FICs.

The FIC's conception, development and implementation can be achieved by a set of operations that will be generically called the *fuzzy-interpolative methodology* FIM. FIM is taking advantage of both linguistic and interpolative nature of the fuzzy systems, combining the advantages of their both sides:

a) The fuzzy sets and fuzzy logic theory will be applied during the conception and the development stages of the control algorithms;

b) The linear interpolations based methods will ensure the implementation stage.

The steps of the fuzzy-interpolative methodology are the following:

a1) the identification of the control solution;

a2) the building of the control rule base of the corresponding fuzzy expert system, represented by McVicar-Whelan tables, in a linguistic manner;

a3) the designing of the Sugeno controller with triangular fuzzy partitions, prod-sum inference and COG defuzzyfication, equivalent to the fuzzy expert system;

b1) the designing of the corresponding look-up table with linear interpolations;

b2) the implementation of the look-up table;

3 The Fuzzy Self Adaptive Interpolative Controllers

J.J. Buckley launched the paradigm of the *universal controller* that could control a wide class of processes without any manual adjustments. Aiming to approach such an ideal structure, the family of *fuzzy self adaptive interpolative controllers* FSAIC presented in figure 2 was introduced in references [4] and [5].

FSAIC has a variable structure. During transient regimes the main controller is a PD one (a 2D look-up-table). Its control surface is almost plane, in order to avoid the distortion of the phase trajectory of the error. During the steady regime an integrative effect is gradually introduced, the structure becoming a PID one. This functionality is achieved with a 3D look-up table having as inputs the control error ε, its derivate ε' and its integrative $\int_\varepsilon$. The different PD tables that are creating the $\int_\varepsilon$ dimension differ only at the central rule, that is activated when ε=zero and ε=zero. Thus the integrative effect is gradually activated, only when steady regimes occur. This controller is called *plane surface adaptive interpolative controller* PSAIC.

The adaptive feature that is creating the FSAIC is introduced by a PD FIC corrector that is acting by mean of a multiplicative correction factor *Gain*.

What is important for our issue is that the design of the adaptive corrector includes a set of *general knowledge on linear PID controllers' adjustment and on linear systems' stability.*

The FSAIC operation relies on the qualitative analyze of the phase trajectory of the error. The strategy is to push the phase trajectory towards ε=zero and ε'=zero point, in a sliding mode like manner. The tactic is to maintain the

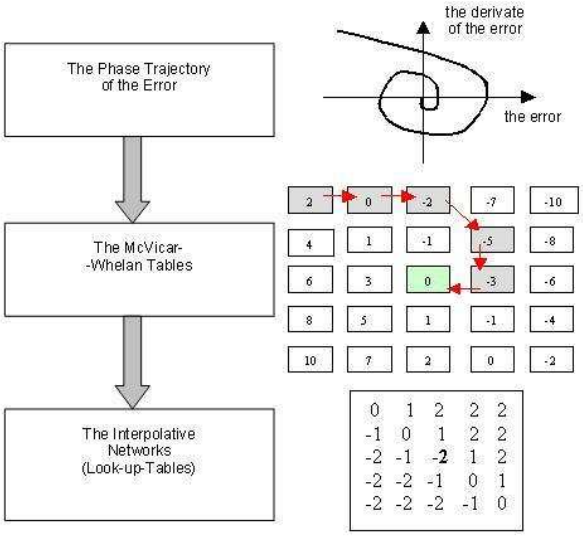


Figure 1: The theoretical tools that are supporting the fuzzy-interpolative methodology

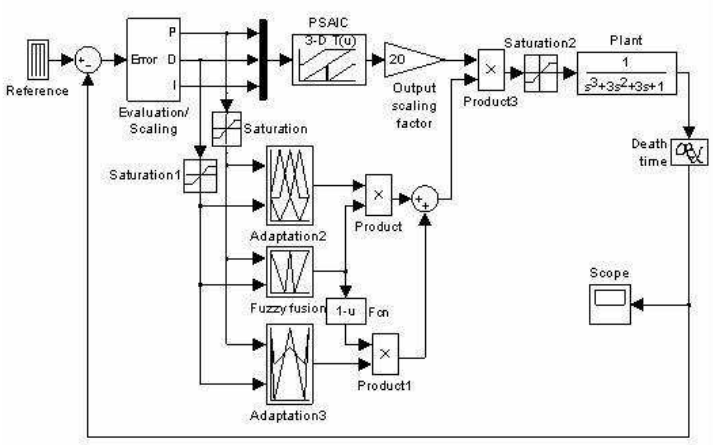


Figure 2: The Fusioned Fuzzy Self Adaptive Interpolative Controller

phase trajectory into quadrants II or IV as long as possible and to avoid quadrants I and III. This task is performed by PSAIC that also needs adaptive capabilities in the case of highly nonlinear or time varying plants. The relevant operating regimes that might occur (transient, steady, oscillating and unstable) need specific adjustments of the controller. Their online identification can be performed by detecting the activation of the rules that are the most relevant signatures of each regime. Thanks to the tabular organization of the rule base (McVicar-Whelan) this operation is simple and can be related to the phase trajectory of the error, as in figure 1.

The most relevant situations, linguistically described, are the following:

1. if ε=zero and ε'=zero the regime is *steady* and the gain is *great*
   This rule is identifying the installation of the steady regime and its effect is to increase the action of PSAIC. This way the control precision is increasing, as well as the sensitivity, for the best possible rejection of the minor perturbations.
2. if ε=zero and ε'=medium or ε=medium and ε'=zero the regime is *transitory* or *oscillatory* and the gain is *medium*
   This situation may appear either in oscillatory regimes or when overshoots are producing, in both situations the gain must be decreased.
3. if sign($\varepsilon \cdot \varepsilon' > 0$) the regime is *unstable* and the gain is *small*
   The system is now firmly installed into quadrants I or III and the best measure against this situation is to reduce the gain at a minimum value, according to the Nyquist stability criterion.

This way system theory knowledge on identification, correction and stability can be embedded, constituting essential pieces of WK for the closed loop controllers. According to our experience adaptive nonlinear PID controllers can cope to almost any technical application if provided with self-adaptive algorithms with relevant WK. An interesting conclusion of this approach is that *instead of developing new control laws we should better concentrate on how WK can be embed into nonlinear PID controllers and to select the relevant pieces of knowledge that are worse to be considered.*

The FIC's capability to embed and to process knowledge is explained by the expert system side of any fuzzy system. As detailed in [10] and some related papers, using FIM in the expert system design generates *fuzzy-interpolative expert systems*, able to cope with the linguistic nature of WK.

4 The Planned Fuzzy Interpolative Controllers

Besides the fundamental system theory knowledge that is governing all the control theory and that is useful for any controller, each application has its own features that are personalizing it. Keeping in sight the specific details of an application may make the difference between success and failure.

Trying to find techniques that are compatible with FIC and allow the representation of the specific WN concerning the controlled process, we have so far considered the *internal models* and the *planners*.

The internal functional models would be the ideal solution for this problem, but unfortunately their implementation in the industrial online control is yet very difficult because of the high computational demands. On the other hand, the planned systems in the sense of ref. [11], which are also solving the specific knowledge handling problem, may be fully compatible with FICs if realizable with mappings or look-up-tables.

The planning systems can harmonize a controller to each specific process if the planners' design is assisted by *functional computer models of the processes*.

The structure of a world knowledge embedding planned controller is presented in figure 3. The WN fuzzy-interpolative expert system is common to all the applications while the planner is personalized. Although the nonlinear PID controller could be realized in any possible technology, the fuzzy-interpolative is the first choice, in order to perfectly match the adaptive part.

Some WK topics that were used so far by our team in different works are: technical data about vehicles, sensors, psychological behavior, physiology (humans and plants), etc. A case study on the traffic management by constant time to collision planners is presented in our joint paper.

5 Conclusions

The world knowledge represents a fundamental resource for the future close loop controllers. If embedded into control algorithms, the general knowledge on the system theory as well as the specific knowledge on the controlled

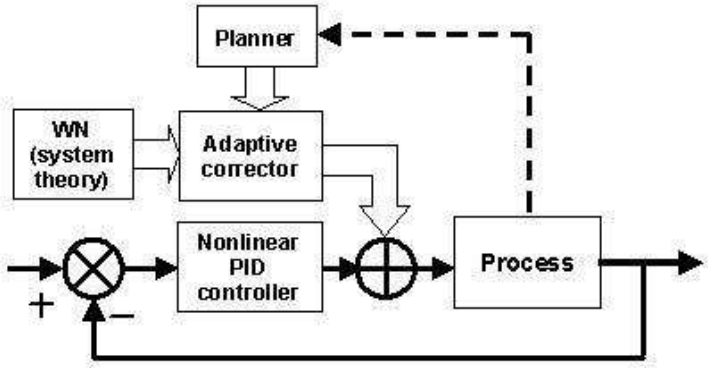


Figure 3: The world knowledge embedding planned fuzzy-interpolative controller

process is able to significantly improve the control performance in any possible sense (precision, robustness, speed, smoothness, etc.) A fundamental theoretical and applicative tool that enables us to provide low level computing devices - μCs, DSPs, etc. with WK is the planned fuzzy-interpolative controller.

References

[1] Lotfi A. Zadeh. From Search Engines to Question-Answering Systems - The Problems of World Knowledge, Relevance, Deduction and Precisiation, *Invited Conference, SOFA'05*, Arad, Romania, August 30, 2005.

[2] L.T. Kóczy, K. Hirota, Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. *Information Sciences*, no. 71, pp. 169-201, 1993.

[3] L.T. Kóczy, K. Hirota, Approximate reasoning by linear rule interpolation and general approximation. *International Journal of Approximate Reasoning*, no. 9, pp. 197-225, 1993.

[4] M. Bălaş. *Regulatoare fuzzy interpolative*. Politehnica Eds., Timisoara, 2002.

[5] M. Bălaş, V. Bălaş. The Family of Self Adaptive Interpolative Controllers. *Proc. of Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'04*, Perugia, July, 2004, pp. 2119-2124.

[6] L.T. Kóczy, M.M. Bălaş, M. Ciugudean, V.E. Bălaş, J. Botzheim. On the Interpolative Side of the Fuzzy Sets. *Proc. of the IEEE International Workshop on Soft Computing Applications SOFA'05*, Szeged-Arad, 27-30 Aug. 2005, pp. 17-23.

[7] M. Bălaş, V. Bălaş. World Knowlwdge for Control Applications. *Proc. of 11th International Conference on Intelligent Engineering Systems INES 2007*, Budapest, 29 June - 1 July, 2007, pp. 225-228.

[8] M. Bălaş, V. Bălaş. Another Possible Way towards the Intelligent Nano-Systems the Fuzzy-Interpolative. *Proc. of Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'06*, Paris, July, 2006, 2135-2141.

[9] V. Bălaş, M. Bălaş. Observing Control Systems by Phase Trajectory of the Error. *WSEAS Transactions on Systems*. Issue 7, vol. 5, July 2006. pp. 1717-1722.

[10] M. Balas. Le flou-interpolatif, present et perspectives. *Seminaire LSIS* St.-Jerome, Marseille, France, 21 sept. 2006.

[11] K.M. Passino, P.J. Antsaklis. Modeling and Analysis of Artificially Intelligent Planning Systems. *Introduction to Intelligent and Autonomous Control*, by P.J. Antsaklis and K.M. Passino, Eds., Kluwer, 1993, pp. 191-214.

[12] M. Balas, V. Balas, J. Duplaix. Optimizing the Distance-Gap between Cars by Constant Time to Collision Planning. *Proc. of IEEE International Symposium on Industrial Electronics ISIE 2007*, June 2007, Vigo, pp. 304-309.

Marius M. Bălaş, Valentina E. Bălaş
Aurel Vlaicu University of Arad, Romania
E-mail: balas@inext.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 33-39

Constant Time to Collision Platoons

Parallel session invited paper

Valentina E. Bălaş, Marius M. Bălaş

Abstract: The paper is presenting a new method for the management of the traffic flow on highways, based on the constant time to collision criterion. The criterion is applied for each car implied in traffic, and for the whole highway. Each car is provided with a constant time to collision cruise controller, which is maintaining optimal distance-gaps between cars, adapted to the speed and to the technical data of the cars. The traffic management center has the possibility to impose the same time to collision over the entire highway. This way the traffic is organizing itself, by distributing the cars such way that the collision risk is uniformly distributed. Simulations are illustrating how the cars are behaving when they are forming highway platoons and how the traffic flow may be controlled by imposing the time to collision.
Keywords: knowledge embedding by computer models, constant time to collision, fuzzy-interpolative cruise controllers.

1 Introduction

Automate driving is enhancing the driving performance and reducing the crash risks. The Advanced Driver Assistance Systems ADAS are such systems [1], [2]. Some of these systems can be linked to cruise control system, allowing the vehicle to slow when catching up the vehicle in front and accelerate again to the preset speed when the traffic allows. A key problem in this issue is the control of the distance gap between cars. In some previous papers [3], [4], [5], we introduced a fuzzy-interpolative distance-gap control method that is using a Constant Time to Collision Planner CTCP, in the sense of the Planning System concept [6]. This approach was also discussed in ref. [7]. The present work is continuing the investigation on the Constant Time to Collision criterion CTTC in the domain of the traffic management. A model of a CTTC platoon is introduced. The simulations are focused on the way in which the cars are forming platoons under the effect of the imposed time to collision, and on the relationship between the imposed CTTC and the traffic intensity.

2 The Constant Time to Collision Criterion

Several indicators measure the characteristics of the traffic flow: the Time-to-Colli-sion TTC, the Time-to-Accident, the Post-Encroachment-Time, the Deceleration-to-Safety-Time, the Number of Shockwaves, etc. [1], [2]. TTC is the time before two following cars (Car2 is following Car1) are colliding, assuming unchanged speeds of both vehicles:

$$TTC = \frac{d}{v_2 - v_1} \tag{1}$$

TTC is linked to the longitudinal driving task. Negative TTC implies that Car1 drives faster, i.e. there is no danger, while positive TTC is leading to unsafe situations. By assessing TTC values at regular time steps or in continuous time, a TTC trajectory of a vehicle can be determined. Doing this for all vehicles present on a road segment one can determine the frequency of the occurrence of certain TTC values, and by comparing these distributions for different scenarios, one can appreciate the traffic safety [2].

The central issue in cars' safety is to impose an appropriate distance between cars, di. The Autonomous Intelligent Cruise Control AICC is imposing a particular polynomial $d_i(v_2)$ law:

$$d_i(v_2) = z_0 + z_1 \cdot v_2 + z_2 \cdot v_2^2 = 3 + z_1 \cdot v_2 + 0.01 \cdot v_2^2 \tag{2}$$

Several settings are recommended, for example $z_1 = 0.8s$ or $z_1 = 0.6s$. Two objections can be drawn against this polynomial $d_i(v_2)$ law:

- no effective adaptation to the traffic intensity is offered: if (3) is tuned for intense traffic, when the traffic is decreasing, the following cars will continue to maintain the same short distance-gaps between them. The driving rules used on highways today are even weaker: "keep distance above 100m" for instance.

- z_1 and z_2 are artificially introduced parameters, they have no significance for humans - highway operators or drivers - and they are not linked to the physical features of the system.

The *Constant Time to Collision* criterion CTTC consists in imposing stabilized TTCs by means of the Car2 cruise controller. Applying CTTC brings two obvious advantages:

- a constant collision risk for each vehicle involved;

- the possibility to control the traffic flow on extended road sections, if each vehicle will apply the same TTC that is currently recommended by the Traffic Management Center [14]: a long TTC means *low traffic flow* and *higher safety* while a short TTC means high traffic flow and higher risk.

The on-line TTC control is not convenient because when the two cars have the same speed the denominator of TTC is turning null: $v_2 - v_1 = 0$. That is why CTTC must be implemented off-line, with the help of $d_i(v_2)$ mappings (fig. 1). The CTTC implementation by $d_i(v_2)$ distance-gap planners is possible because *a distance gap planner using* TTC *will produce* CTTC. We studied this method by computer simulations, using a Matlab-Simulink model of the tandem Car1-Car2, introduced in other previous papers [3], [4], [5], [9], [13].

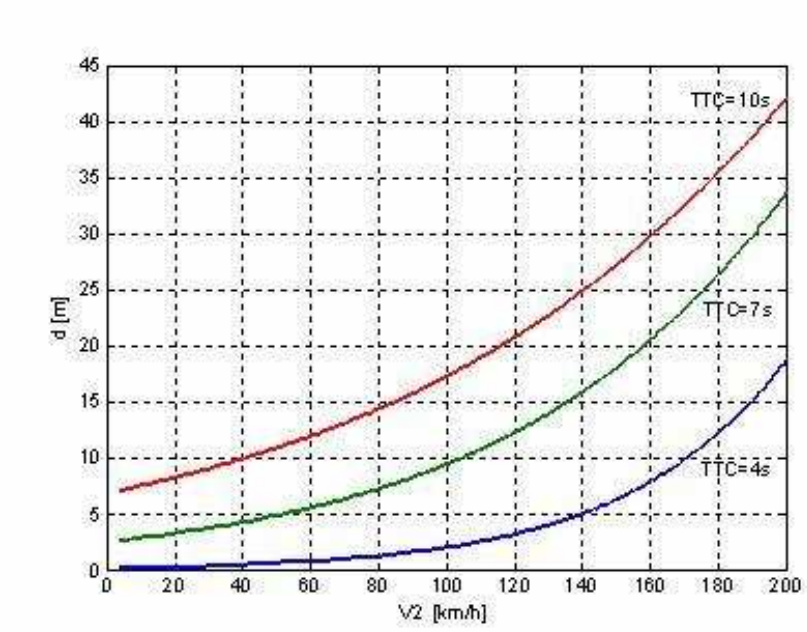


Figure 1: The recorded $d_i(v_2)$ mappings for three different TTC

Since the design of the planners is performed with the help of functional models of the cars, accurate knowledge about the specific behavior and parameters of each car (traction and braking forces, weight, aerodynamic coefficient, etc.) can be taken into account, which is not possible to the simplified and leveling analytic model (2).

The application of this method is imposing to the car manufacturers to provide each type of automobile with a computer model.

The distance-gap planners are designed as follows. The simulation scenario consists in braking Car1 until the car is immobilized, starting from a high initial speed. A TTC controller is driving the Car2 traction/braking force such way that during the whole simulation TTC is stabilized to a desired constant value. The continuous braking allow us to avoid the $v_2 - v_1 = 0$ case. We will use the recorded d mapping as the desired $d_i(v_2)$ planner for the given TTC. The figure 1 planners are determined for three TTC values: 4s, 7s and 10s. These planners can be easily implemented with the help of the look-up tables with linear interpolation.

The use of the CTTC planning technique is essentially facilitating the task of the distance controller that is actually driving the traction/braking force of a real car during the cruise regime, as shown in fig. 2. Very simple fuzzy-interpolative PD controllers or even linear controllers can such way cope with the car following task [4].

The implementations can be basically achieved by look-up-table techniques.

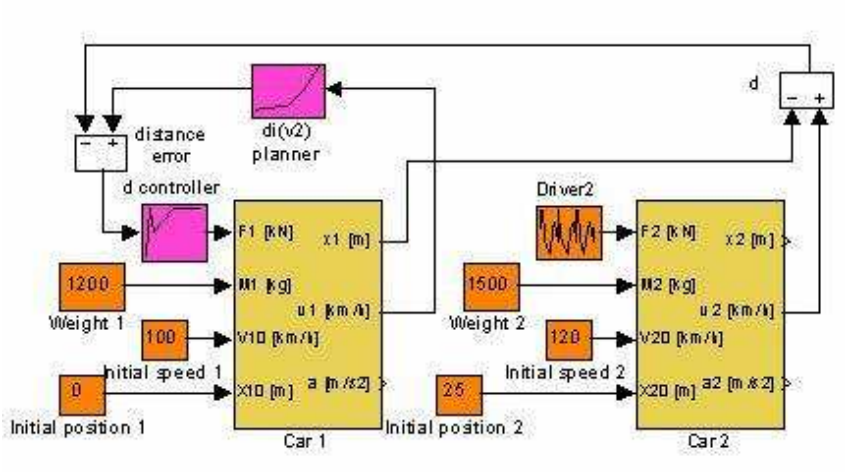


Figure 2: A cruise control system with distance controller and CTTC $d_i(v_2)$ planner

3 The Traffic Management by Constant Time to Collision

The superior application level for the CTTC criterion is the management of the traffic over extended highways' segments. Assuming that each car is provided with a cruise controller with CTTC planner, the Traffic Management Center TMC has the possibility to impose the same TTC to all the cars. This way the highway system becomes a distributed one. Each car is trying to reach and to maintain the position that respects the imposed TTC to the previous car. This trend has as major advantage a constant distribution of the collision risk for each car.

Lets consider that TMC is imposing a 7s TTC. If the traffic is not too intense, the tendency of the cars will be to form platoons that are able to maintain TTC=7s. It is to remark that the distance between the cars belonging to the same platoon are not necessarily identical, even for constant speeds, because each type of cars has its particular $d_i(v_2)$ planner, in accordance to its technical parameters (weight, aerodynamics, engine power, brakes, etc.) If the traffic is beginning to decrease the number of the cars that are included into platoons will decrease too, and empty zones will develop on the highway. In this case TMC should increase the imposed TMC value, either continuously or by discrete values, say TMC=10s. This way the disposable space of the highway will be better covered and the collision risk will decrease for each car.

In the opposite case, if the traffic is increasing, the cars will not be able to maintain the desired TTC and the corresponding distance-gaps. TMC will be forced to reduce the imposed TTC, either continuously or by discrete values, say TMC=4s. This way the density of the traffic will increase and the collision risk will increase too, but this will happen in a smooth and controlled manner, the risk continuing to be equally distributed over each car.

Our research on this matter is only at the initial stage, but the preliminary simulations are confirming that CTTC criterion is potentially able to cope with the highway traffic.

4 A CTTC platoon model

The following Simulink-Matlab model allows us to simulate on computer the behavior of the CTTC highway. As showed in fig. 3, the model is addressed to a five car group.

Each car has its own technical parameters: weights between 1000 and 1400kg (see variables M), engine powers between 100% and 180% of the generic Car1 power (see variables Gain) and its own CTTC planner (see the Di(V2,TTC) look-up-tables). The initial speed Vo and position Xo of each car can be as well adjusted.

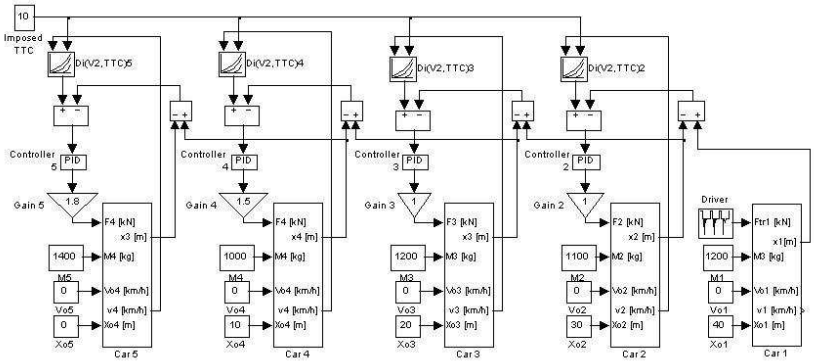


Figure 3: The five cars TTC platoon Simulink-Matlab model

The model is offering the time variation of the aimed parameters: speeds and positions of each car, distance-gaps between cars, the length of the platoon, etc.

5 Simulation results

The scenario of the simulations, imposed by the Car1's driver, is including a 66s acceleration from 0 to 200km/h, a steady floor at 200km/h until 100s, a 20s deceleration to 135/h, another floor at 135km/h, followed by a second deceleration of about 40s that is positioning Car1 at a steady 100km/h.

The first simulation, presented in fig. 4, is illustrating the global behavior of the five car TTC platoon, for TTC=7s.

One can also observe the continuous variation of the platoon's length with the speed, detailed in fig. 5.

In the fig. 6 simulation, executed for TTC=15s, one details the formation of the platoon. One can observe the behavior of the five cars that are starting from random initial positions and are forming the platoon in less than 10s.

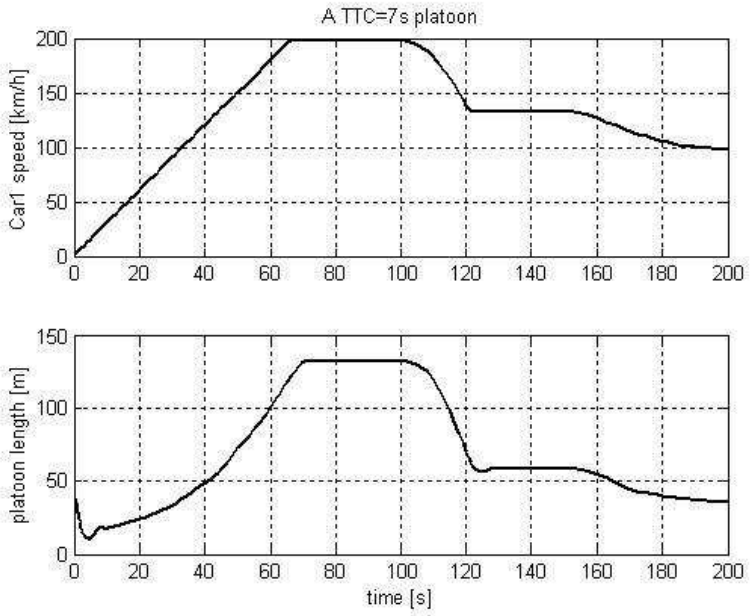


Figure 4: A CTTC platoon simulation, with Car1's speed and the platoon's length

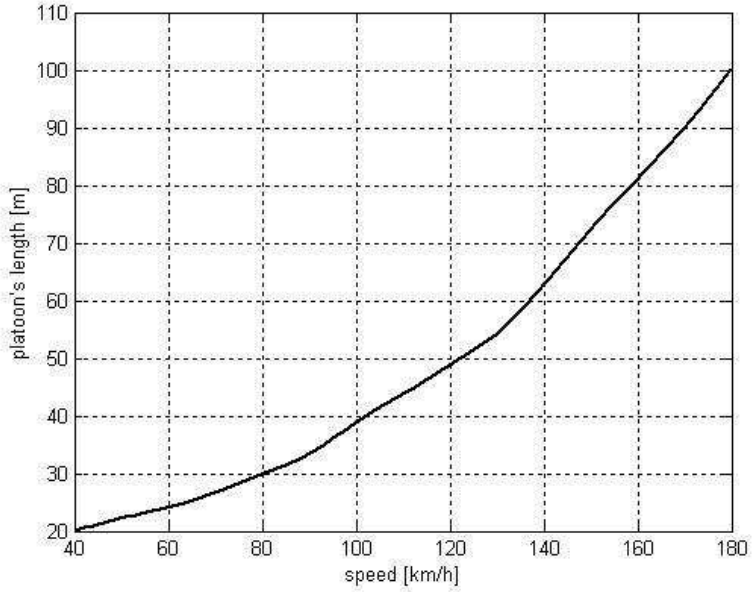


Figure 5: The platoon's length dependence with the speed, for TTC=7s

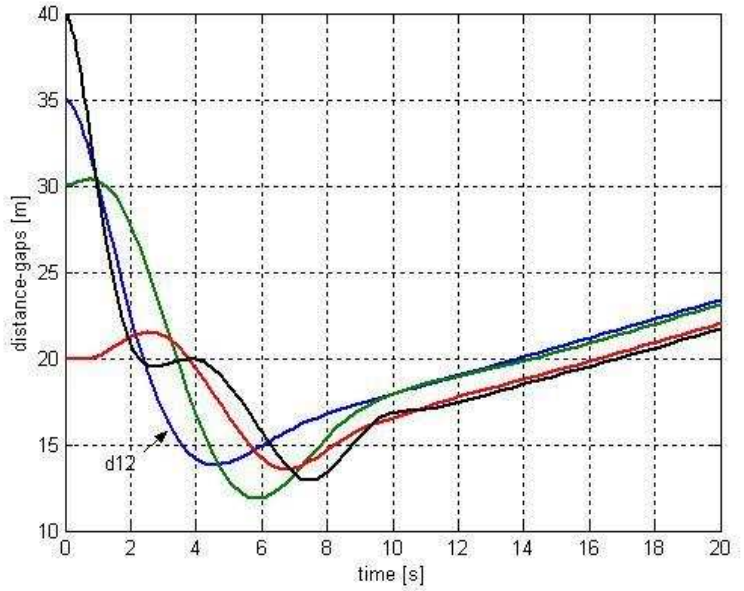


Figure 6: The platoon's aggregation

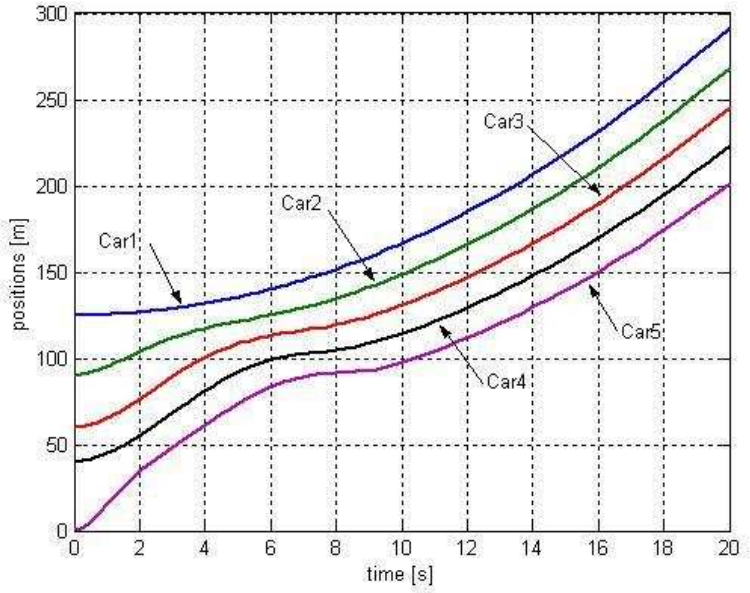


Figure 7: The positions of the cars during the platoon's aggregation

6 Conclusions

The time to collision criterion can be used in the highway traffic management. If each car is provided with a constant time to collision cruise controller, the traffic management center can impose the same time to collision to all the cars. Such way the high-way system becomes a distributed one, each car trying to reach and to maintain the position that respects the imposed time to collision to the previous car. The method keeps constant the collision risk for over all the cars of the highway. Besides the simplicity and the advantageous interpolative implementation, all the time to collision based tools have a common feature: they are embedding precise knowledge about the technical data of the automobiles thanks to the functional computer model that stands behind their design. This adaptive capability is promising to improve the future highway traffic that is presenting so many elements of uncertainty.

References

[1] A.R. Girard, J. Borges de Sousa, J.A. Misener and J. K. Hendrick. *A Control Architecture for Integrated Cooperative Cruise Control and Collision Warning Systems.* Berkeley University of California, http://path. berkeley.edu/ anouck/papers/cdc01inv3102.pdf.

[2] M.M. Minderhoud and S.P. Hoogendoorn. *Extended Time-to-Collision Safety Measures for ADAS Safety Assessment*, Delft University of Technology, http://www. Delft 2001.tudelft. nl/paper %20 files/ paper1145.doc

[3] M. Balas, C. Barna. Using CCD cameras for the car following algorithms. *Proc. of IEEE International Symposium on Industrial Electronics* ISIE'05, Dubrovnik, 20-23 June, 2004, pp. 57-62.

[4] M. Balas, V. Balas. Optimizing the Distance-Gap between Cars by Fuzzy-Interpolative Control with Time to Collision Planning. *Proc. of The IEEE International Conference on Mechatronics*, Budapest, Hungary ICM'06 Budapest, 3-5 July, 2006, pp. 215-218.

[5] V. Balas, M. Balas. Driver assisting by Inverse Time to Collision. *Proc. of The IEEE International Conference on Mechatronics*, Budapest, Hungary WAC'06 Budapest, 24-27 July, 2006.

[6] K.M. Passino, P.J. Antsaklis. Modeling and Analysis of Artificially Intelligent Planning Systems. *Introduction to Intelligent and Autonomous Control*, by P.J. Antsaklis and K.M. Passino, Eds., Kluwer, 1993, pp. 191-214.

[7] M. Balas. Le flou-interpolatif, present et perspectives. *Seminaire LSIS St.-Jerome*, Marseille, France, 21 sept. 2006.

[8] M. Balas. *Regulatoare fuzzy interpolative*. Editura Politehnica Timisoara, 2002.

[9] L.T. Kóczy, M.M. Balas, M. Ciugudean, V.E. Balas, J. Botzheim. On the Interpolative Side of the Fuzzy Sets. *Proc. of the IEEE International Workshop on Soft Computing Applications* SOFA'05, Szeged-Arad, 27-30 Aug. 2005, pp. 17-23.

[10] R.E. Precup, S. Preitl, M. Balas, V. Balas. Fuzzy Controllers for Tire Slip Control in Anti-Lock Braking Systems. *Proc. of The IEEE International Conference on Fuzzy Systems* FUZZ-IEEE'04, Budapest, 25-29 July, 2004, pp. 1317-1322.

[11] M. Balas. World knowledge for controllers. *International Symposium Research and Education in Innovation Era*, Arad, 16-18 Nov. 2006, Section 3, pg. 531-534.

[12] Y. Zhang, E.B. Kosmatopoulos, P.A. Ioannou, C.C. Chien. Autonomous Intelligent Cruise Control Using Front and Back Information for Tight Vehicle Following Maneuvers. *IEEE Trans. on Vehicular Technology*, vol. 48, no. 1, January 1999, pp. 319-328.

[13] M. Balas, V. Balas, J. Duplaix. Optimizing the Distance-Gap between Cars by Constant Time to Collision Planning. *Proc. of IEEE International Symposium on Industrial Electronics* ISIE 2007, June 2007, Vigo, pp. 304-309.

[14] ITS Decision. *Traffic Management Centers* http://www.calccit.org/itsdecision/serv_and_tech/ Traffic_management/TMC/tmc_summary.html

Marius M. Balas, Valentina E. Balas
Aurel Vlaicu University of Arad, Romania
E-mail: balas@inext.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 40-54

E-Maieutics. Rationale and Approach

Parallel session invited paper

Boldur E. Bărbat

It is a miracle that curiosity survives formal education.
ALBERT EINSTEIN

Abstract: Asserting that customary e-Learning is outdated in the era of "computing as interaction" by the context (uncertain, rapidly changing environments), targets (personalised dynamic knowledge), nature (non-algorithmic information processing), and methods (andragogy tending towards heutagogy), the paper aims at introducing the concept of e-*Maieutics* as alternative to conventional e-Learning and at illustrating it in experimental models, where maieutics is action-oriented and highly personalised, while "e-" is carried out through virtual entities interacting with the learner as interface agents. A next target is to ease a paradigmatic shift in artificial intelligence as a whole, by providing an affordable test-bench for two innovative agent features (developed within other research tasks): protensity-based Computer-Aided Semiosis and bodiless agent self-awareness. (This target is not explicitly dealt with here.) The aim is split into three objectives: a) offering a convincing rationale for the new concept; b) outlining an agent-oriented (meta-)approach to its application in open, dynamic, and uncertain environments; c) profiling the main design-space dimensions for e-Maieutic applications (life-long learning, bounded rationality, affective computing, ethical behaviour), as substitute for a plausibility probe, supplemented with some aspects requested by the test-bench role (mainly, a syncretic sense of time). The paper focuses on the trends that require a paradigmatic shift and support the intended path and reviews very briefly some agent-oriented mechanisms used in implementing the new features. Among the conclusions: e-Maieutics is a viable alternative approach to e-Learning, best suited to dynamic and uncertain environments (hosting nontrivial interactive applications); the outlined design space for Socratic agents opens promising paths for both e-Learning and agent-oriented software; the concepts and mechanisms proposed offer a suitable test-bench for related research.
Keywords: Non-algorithmic e-Learning; open, heterogeneous, dynamic and uncertain environments (OHDUE); e-Maieutics; life-long learning; agent-oriented software.

1 Introduction. A new "e-Catchphrase"?

It is a widespread feeling that:

a) Formal education does not meet its targets - and the motto shows that this belief is old, not confined to losers or to socio-historical environments, and insinuates more than failure, namely harm.

b) e-Learning should be close to a panacea - at least like e-mail was for conventional mail.

c) That it is not - despite forty years of efforts and undeniable improvement.

As regards causes and expectations, the quasi-unanimity stops, spawning an opinion palette ranging from "IT will solve it soon" to "there is no effective 'e-solution' for societal problems; worse, any such solution, involving intensive human-computer interaction, is frustrating, counterproductive or illusive". Such paradoxical divergence suggests that both *ends* (teaching strategies) and *means* ("e-" approaches) should be revisited. (Below they are revisited, however, not using traditional "Means-Ends Analysis", but from a stance in line with a new paradigm.)

The paper asserts that a *paradigm shift* (in the sense of Kuhn) is needed regarding both *learning* and "*e-*" and, starting from here, aims at explaining and endorsing the worth of admitting that learning should be *action-oriented* (i.e., promoting rather dynamic than static knowledge) and highly personalised, while "*e-*" should be

agent-oriented (i.e., carried out through virtual entities interacting with the learner as interface agents). A second (but not secondary) target is to ease a paradigmatic shift in artificial intelligence (AI) as a whole too, by providing an affordable test-bench for two innovative features (aimed at within two related, long-range, agent-oriented undertakings): protensity-based Computer-Aided Semiosis (CAS) [13] [36] (details in a related paper, presented at this conference) and bodiless agent self-awareness [11, 8, 22, 23] . (Since this target is pursued in [6], to impair redundancy, it is not explicitly dealt with here.)

To give a convincingly negative answer to the question in the section title, a research undertaking should provide evidence based on several applications validated in vivo. This paper confines itself to present a careful rationale, and an affordable approach, because:

a) It is linked to computer science endeavours hosting e-Maieutic applications, as well as applications with specific maieutic features, transcending the common denominator presented here.

b) Since two of them are ongoing projects in innovative artificial intelligence (AI) sub-domains, they resolve so far toy problems.

c) Even for real-world problems they do not go beyond the stage of experimental models.

d) They are described in more context-relevant related papers (details in [6]).

Though, since no approach can convince prior to - at least - a qualitative validation, but such validation is impracticable before some experimental models are operational, the paper has to go beyond its title and initial objectives, offering also some substitute for a plausibility probe. Thus, the main e-Maieutic design-space dimensions suggesting the directions for developing *e-maieuts* are given, together with brief information about agent-oriented mechanisms used in implementing the new features. Those mechanisms are described in detail in [3, 12, 11, 35, 14, 8], were tested in experimental models for a virtual disc jockey (in the related paper mentioned above) and for a virtual guitar teacher [23] and served to give a rough idea about tailoring a virtual Socratic nurse (*VISON*) [5].

Hence, the aim is to introduce the concept of *e-Maieutics*, to defend it as (essentially nonalgorithmic) alternative to conventional e-Learning, suited to both *content* (specific to life-long learning) and *setting* (dynamic and uncertain environments, hosting most nontrivial interactive applications), and to illustrate it in experimental models, where both *maieutics* and "*e-*" fulfil the requirements stated above. This aim is split into three objectives:

a) providing a convincing rationale for the new concept;

b) outlining an agent-oriented (meta-)approach to its application in dynamic and uncertain environments;

c) profiling the main design-space dimensions for e-Maieutic applications (life-long learning, bounded rationality, affective computing, ethical behaviour), supplemented with some primeval aspects requested by the test-bench role (above all, regarding a syncretic sense of time; details in [6]), as well as with some features specific to the virtual entity they represent (disc jockey, guitar teacher, nurse).

On the other hand, since *e-Maieutics* is coined now, *e-Nursing* is explicitly polysemantic, and *e-Learning* is not in good shape, a prolegomena about the shifting meaning of "*e-* prefixing (almost) everything", seems required. Likewise, since *e-Maieutics* is a new concept and this undertaking as a whole is unconventional, related work *proprio sensu* does not exist. Hence, the usual section succession is reversed: after scrutinising *epistemic shifts*, making "*e-*" a *risky prefix* (Section 2), the *rationale* for e-Maieutics is detailed focusing on its *context and links* and reminding its *roots* (Section 3), and *related work* is presented *filtered* through the sieve of the aim looked for, *focusing on the trends* that require the paradigmatic shift and support the intended path (Section 4). The *(meta-)approach* is proposed as a *two-way line of attack* (Section 5). The specific *design-space dimensions* are commented upon and the mechanisms are very briefly mentioned in Section 6. *Conclusions and future work* (Section 7) close the paper.

2 Epistemic Shifts. Is "e-" a Risky Prefix?

Why starting with epistemological problems? Consider the following working[1] assumption:

[1] Here, "working" has the same pragmatic connotation as in the syntagm "working definition".

To participate efficiently in something, individuals have to be rather *proactive* (i.e., *pushing* events) than *reactive* (i.e., *pulled* by them). Taking into account the targets of learning this assertion could be considered quite an axiom. Thus, it can be regarded as the first link of an implication chain: *proactiveness* requires a more refined approach than *reactivity*; in complex settings a refined approach involves a *transdisciplinary coherent perspective*; to arrive at such a perspective, particular stances must be brought to a *common denominator*; that denominator involves reconciling - or at least isolating - different *mindsets*; this entails an epistemological problem (that of understanding unambiguously the concepts involved). To get closer, "*something*" is instantiated to "*Internet era*" and all implications are formalised through "→"; the (shortened) deduction chain becomes now:

- Efficient participation in the Internet era → user-pushed information and communication technologies (ICTs; here this less familiar term is used only to stress the scope of the process; focus is on IT).

- User-pushed technology → transdisciplinary perspective.

- Transdisciplinarity → common denominator.

- Common denominator → reconciling/isolating mindsets.

- Isolating mindsets → evaluating concepts.

Since the target of reconciling mindsets is yet too far - and is a moving one - the section aims at identifying, and investigating some of those mindsets together with the different connotations of common concepts, biased by deep rooted attitudes. Myths are always deep rooted: for instance, the myth of "inexorable technological determinism" (never truly dead but revived by every new powerful technology). This myth is expressed in several varieties; its most radical form is: "modern technology is (or, at least, show facets that are) intrinsically evil, endangering eternal human values and thus the human species; hence, it should be avoided or, if that is impossible, at least denounced"; one of the most moderate variant is "modern technologies are unavoidable despite their evident, very frustrating side effects; hence, to defend human values, such side effects must be denounced and, whenever possible, new methods should be used only when they can be simply applied to conventional approaches". Only the more moderate and easy to defend posture will be considered (thus, the argument "don't listen to extremists" will be unsuitable).

As any *Zeitgeist*-component, such myths engender a treacherous corollary: widening the gap between teachers, learners and IT designers, they impair transdisciplinarity. In addition, understanding "in transdisciplinary manner" means being aware of all undertones (moreover, since some of them are elusive or even subconscious). Because this is impossible to reach within the scope of a single section, the idea is to propose rather a *catalyst* than a *glossary*, focusing on diverging perspectives and on their consequences regarding the paths to follow for more effective e-Learning.

Looking at transdisciplinarity from a semiotic point of view, the focus should be on *semantics* and on *pragmatics* at least as much as each educational culture has focused on *syntactics*. In other words, the relationship between (similar) signs and (different) meanings and the relationship between those signs and (very diverse and very many) stakeholders involved in e-Learning must be explored more accurately.

The discrepancies regarding the connotations of concepts are more visible than those of mindsets because of five reasons:

a) they derive from the general perspective and often develop it further;

b) every mindset express itself through a set of concepts;

c) since mindsets are sometimes vague or subconscious, they are mostly implicit and can stay hidden, whereas concepts have to be dealt with, hence must be explicit;

d) the resulting confusion is more obvious;

e) partial and ambiguous synonymy.

Since a detailed debate would be ineffectual until the mindsets are isolated, the divergent connotations are illustrated by the most pertinent example of "e-" (often written without the hyphen).

This widespread prefix started as an abbreviation for "*electronic*". Nowadays, it may be attached to anything that has moved from a traditional form to its IT[2] alternative (e.g., e-mail, e-commerce, e-business, e-learning, or e-procurement). In other words and maybe oversimplified, "something put up and/or available via the computer and the Internet".

Ambiguity appears when the prefix is used metaphorically. For instance, "The *eEurope 2005* Action Plan was launched at the Seville European Council in June 2002 and endorsed by the Council of Ministers in the *eEurope Resolution* of January 2003" (www.europa.eu.int/information_society/eeurope). Reading the text, the meaning is clear and legitimate but, taking only the term "*eEurope*", it is confusing, since there is not an *alternative*, "*electronic* Europe", as in the case of e-mail.

Seemingly the same - but a big step on the way to augment confusion - is the case with another familiar term: "*E-democracy*, a portmanteau of "electronic" and "democracy", comprises the use of electronic communications technologies, such as the Internet, in enhancing democratic processes within a democratic republic or representative democracy. [...] The term is both descriptive and prescriptive. [...] E-democracy is also sometimes referred to as *cyberdemocracy* or *digital democracy*. Prior to 1994, when the term e-democracy was coined [...], the term *teledemocracy* was prevalent. [...] "Teledemocracy" then is an umbrella term that includes both "e-democracy", "deliberative democracy" and many types of "direct democracy"" (http://en.wikipedia.org/wiki/E-democracy).

In fact, there are two treacherous paths in this balanced, encyclopaedia-style, description:

a) *The undesirable prescriptive character*. Since "*prescriptive concept*" is not defined explicitly, to pass up the common connotations, the term will be presented from a technological perspective (an approach in managing ship structural development): "The prescriptive concept is universally understood. In childhood, parents prescribe the code of behaviour, later the school teacher has a similar role, at college the rules of the institution must be kept if the course is to be successfully completed, and the majority of adults recognise that the laws of the land have to be kept. It could be said, in fact, that the prescriptive concept is ingrained in our thinking" (www.shipstructure.org/sss2000/Kuo_7.pdf -). To circumvent inflated mindsets, here follows a definition from a neutral area (grammatical correctness): "*prescription* is the laying down or *prescribing* of normative rules for the use of a language, or the making of recommendations for effective language usage" (http://en.wikipedia.org/wiki/Prescriptive). Mutatis mutandis, who intends to *prescribe* normative rules for the use of "any kind of democracy"? Of course, it is not forbidden, but why prescribing rules by the means of a *concept*, expressed through a *prefix*, in an ambiguous context of a debatable domain? Besides doubts about prescriptive concepts (after such a "prescriptive" XXth century) in any field of interests, using them linked to technology is harmful, no matter the stance involved: technological *dictatorship* ("democracy *must* become e-democracy in line with the prescriptions") or technological *determinism* ("just another example how evil can become those imposed e-technologies"!).

b) *The risks of "Traduttore-traditore"*. Paradoxically, the risks are higher when many *seeming* synonyms try to help. For instance, the same encyclopaedia uses (fortunately, this time) the term "electronic direct democracy" (http://en.wikipedia.org/wiki/Direct_democracy), being at odds with the poor umbrella term mentioned above.

However, when using "*digital*" instead of "*e-*" things get both worse and revealing. The first technological connotations were those of "*discrete*". Later in its semantic journey, "*digital*" describes electronic technology that encodes information in binary form. Since computers work in both *electronic* and *digital* form, the terms are quasi-synonyms in IT ("*digital*" is more accurate semantically); thus, old terms stay with "*e-*" (e.g., *e-mail*), while new ones use "*digital*" (e.g., *digital divide*). Nevertheless, a subconscious influence of an "anti-technology mindset" is yet hidden: whereas nature - humans included - was always intrinsically *analog*, the invading IT was *digital* and, until a decade ago, was unable to interact with its users through an entirely analog interface. Albeit IT can afford now interfaces enabling users to interact with in their ancestral, analog manner, the nearly forty years of manifest digital IT structure, induced the feeling that "*digital*" involves a dangerous feature (the next syntagm pair is enlightening: "*electronic processing*" versus "*digital manipulation*"). As a result, bad circumstances get as attribute "*digital*", while better ones are seen as "*electronic*". Confusion reaches its pinnacle when old technologies progressed from an analog structure to a digital one. Now, "*digital*" is used again reasonably in "*digital camera*" (because it replaced the old one, based on analog technology) and acceptable in "*digital television*" (same reason).

[2]In its prehistory (i.e., in the forgotten era when mainframes were kings), the technology that was called (some years later) in Europe "*Informatics*", was known as "*EDP*" (*Electronic Data Processing*). Even in Europe the French term "*informatique*" was adopted reluctantly at the beginning (for instance, in Germany, many used the old term: Elektronische Datenverarbeitung). Does it matter? Yes, since familiar terms express the Zeitgeist.

"Acceptable" tries here to show the wrong focus: this new *digital* television has such a great user impact because it is *interactive*, not because it is *digital*. This is an example of substituting "*What?*" (the *architecture*: interactivity is an innovative feature) by "*How?*" (the *structure*: the irrelevant digital infrastructure). On the other hand, "*digital*" is ridiculous in "*digital privacy*" and mystifying in "*digital ethics*" (the matter is dealt with in detail in [10]).

In short, the meaning of "*e-*" should shift paradigmatically from denoting a *technological means* (as in "*e-mail*") towards a *socio-cultural context* (as in "*eEurope*"). Otherwise, it would rather not matter.

3 Rationale: Context, Links, and Roots

When the concept to be launched involves a paradigm shift, the rationale must show why the problem cannot be solved within the still dominant IT paradigm - labelled in [36] ALNUOP (*Alg*orithms, *Nu*merical, *Op*timization) - namely why the new concept (here, e-Maieutics) is required by the new real-world problem (here, e-Learning). In other words, prove that conventional e-Learning is outdated by the *context* (uncertain, rapidly changing environments), *targets* (personalised dynamic knowledge), *nature* (non-algorithmic information processing), and *methods* (andragogy tending towards heutagogy) of learning in the Internet era (of "computing as interaction"). (The unavoidability of non-algorithmic approaches in agent-oriented software was insisted on recently also outside e-Learning requirements, in [14, 7].)

Context. The broad context, described recently in [12, 14, 7], shows that IT environments, except for some trivial applications, are open, heterogeneous, dynamic and uncertain (OHDUE), where both *information* and its *processing rules* are revisable, fuzzy, uncertain and, hence, intrinsically non-deterministic. Except some subdomains were automation is not yet "balanced" (in the sense of BASYS, involving decisive human intervention), deterministic applications either vanish (e.g., expert systems are replaced by agents), or cause almost universal discontent (albeit the reasons are not clearly identified, as in the case of e-Learning).

For e-*Learning*, the context was outlined in [35] and is here abridged. The knowledge-based society entails new targets for e-Learning because:

a) Humans must (inter)act in OHDUE quite different to the way they are familiar with.

b) The challenges to cope with are major and involve other requirements.

c) IT advanced dramatically offering possibilities, means, perspectives, and approaches unthinkable about forty years ago when e-Learning took off.

d) Without an anthropocentric and transdisciplinary approach end-user acceptance will not match the huge technological potential on hand.

As regards the *learning process* as such - prefixed with "*e-*" or not - the viewpoint is that human learning is best described by the information-processing approach in cognitive psychology, in line with the ideas endorsed in [1]: "Most modern information-processing theories are "learning-by-doing" theories which imply that learning would occur best with a combination of abstract instruction and concrete illustrations [...] combining abstract instruction with specific concrete examples [...] is better than either one alone". Learning should be considered - in both humans and agents - as a process where most effectiveness is reached through a blend of symbolic ("left-hemisphere"-like) and subsymbolic ("right-hemisphere"-like) modi operandi. Nowadays, the approach is much closer to "by rote learning". Thus, the balance has to be redressed, favouring right hemisphere tactic.

Moreover, the geometrically increasing computing power (due to Moore's law) entails that "remembering information" is almost not anymore needed, since the computer remembers much better and faster (and WWW almost never forgets). Hence, the focus is on *understanding* (as aim) and on *involving* (as means). Coincidentally (or not?), in the era of "computing as interaction", involving humans seems rather natural. (The agent in the experimental model in [35] could be seen as an oversimplified kind of "personalised and intensive Google".) Although the needed technology is available, as shown in [5], the title of [43] puts the right question: "Dialogue and Discourse: Are We Having the Right Conversations?"

On the contrary, albeit a broad consensus that *learning* (at least, high-level one) is innately non-algorithmic [1, 25, 37], the yet crushing predominance of algorithmic software impairs common approaches to non-algorithmic *teaching*. In addition, non-algorithmic software is disregarded on the whole. Besides, the educational paradigm is still reluctant to accept the shift from obsolete pedagogy, towards *life-long learning* - applying modern *andragogy* and, perhaps, yet syncretic *heutagogy* (see next sections): "While institutions are recognizing the value of online education, perceptions have been slower to change. Unfortunately, [...] poor attempts at online education have

sometimes tended to engulf the learner, causing frustration and defeat. Studies indicate that nearly 85 percent of learners involved in [...] e-learning quit before finishing their program (2002)" [38].

Finally, the main pillar of any rationale: practical usefulness. Most relevant is the case of nursing where maieutics is almost unavoidable: "The nursing process is [...] often supported by nursing models or philosophies" (en.wikipedia.org/wiki/Nursing_process). For nursing a modern and popular model is the "Synergy model" (en.wikipedia.org/wiki/Synergy_model_of_nursing): "designed to pair the needs of the patient and their family with the strengths of the nurse providing care. For instance, if a patient comes from a different culture than the nurse, the nurse who is an expert [...] in the Response to Diversity competency would be able to evaluate her own biases and beliefs [...]. This theory may be perceived as difficult to put into practice, however with the *nursing shortages and tight staffing ratios*[3] ." (Corollary: virtual nurses are no luxury anymore, they are needed and the requirements become higher.)

Links. The strong bonds to other two long-range research projects (mentioned in Section 1) strengthen the immediate rationale sustained above. Both projects need a non-conventional approach to e-Learning (non-algorithmic, action-oriented, and highly personalised), implemented via agent-oriented software engineering) as test-bench for their innovative agent features because:

a) The second stage of the "teleoreactive" and skill oriented "Protensional Agent" (PA) is a guitar teacher [23, 36] guided by andragogic methods (its first form, as improved disk jockey, is presented in a related paper).

b) Self-referencing time-aware agents - or "Hofstadter Agents" (HA) - could prove their pragmatic relevance only within maieutic-like learning settings where they can show a cognitive architecture close to that of the humans they interact with, for at least five reasons:

- **b1)** "The activity of a teacher is relevant to the extent that it causes students to engage in activities they would not otherwise engage in" [1].
- **b2)** Corollary: "trainer and trainee should share a common ontology - at least for basic communication [...] To insert the concept of learning into a common ontology (humans *know* when *they* learned) and to exploit it, the agent must have minimal introspection ability" [35].
- **b3)** Likewise, "given the basic role of time in human cognition and consciousness, agents with a powerful temporal dimension ("thick time" included)" [8] are vital.
- **b4)** "Higher order thinking requires self-regulation; someone else is not giving directions" [37].
- **b5)** Agents should assess themselves through a "Simon-type machine learning" performance metrics (details in [35, 8]).

There is also a long-range reason: investigating the chances that inductive learning could allow e-Maieuts and humans to *learn* from each other to *teach*.

In fact, the need for a strong temporal dimension, inherent to maieutic learning, is vital for both HA and PA. Obvious for PA, this requirement becomes crucial for HA, because it is a purely software agent. Indeed, the "challenge is major: could bodiless agents (lacking any spatial sense or haptic proprioception) be self-aware? Above all, when self-awareness is expected to emerge from "strange loops" à la Hofstadter, emulated via Gödelian self-reference" [8]. If in 2007 it was considered that the strong temporal dimension, besides "its intrinsic architectonic value, it could be helpful in future "pseudosomatoception" as surrogate for the lacking sense of space and haptic proprioception" [11], at the present stage of research it is clear that "the hope is that *space* can be - at least, partially - substituted by *time*" [8]. As a result, the test-bench role of e-Maieutic applications for devising agent time becomes a sine qua non condition for both undertakings. The cardinal issue of "three kinds of time, i.e., physical ("Caesium time", *TCs*), psychological ("Carbon time", *TC*) and agent ("Silicon time", *TSi*)" [8] is dealt with in [6] in the much larger context of natural language computation.

In short, an alternative path to conventional e-Learning - no matter its label - is *needed* because:

a) Moore's law effects are striking: Google becomes a widespread e-tutor (always necessary, sometimes sufficient);

b) learning occurs in OHDUE, where intense non-deterministic interactions decrease rapidly the role of algorithms;

[3] My italics (BB).

c) the emphasis moves from *static* towards *dynamic* knowledge. Moreover, such a path is necessary as test field for the architectonics of advanced agent-oriented software.

Finally, some research directions outside the scope of the paper: "One hallmark of Socratic questioning is that typically there is more than one "correct" answer, and more often, no clear answer at all. [...] The Socratic method has been adapted for psychotherapy, most prominently in Classical Adlerian psychotherapy and Cognitive therapy. It can be used to clarify meaning, feeling, and consequences, as well as to gradually unfold insight, or explore alternative actions" (http://en.wikipedia.org/wiki/Socratic_method).

Roots. Since the history of all agent-oriented research is abridged in [6], and the history regarding e-Nursing in [5], here are reminded only the roots in prehistory (before 2005) comprising research in four interrelated areas:

a) uncertain knowledge processing (mainly asynchronous reaction to environment stimuli and temporal aspects in agents);

b) anthropocentric systems (agent-oriented captology, multimodal analog human-agent interaction, user-driven heuristics);

c) affective computing (pathematic agents designed as virtual therapists, emotion as asymmetric temporal function, controlling ethical agent behaviour);

d) mechanisms for agent reactivity (e.g., clone-based polymorphism, exception handling).

Approaches, results, and standpoint are presented in [3, 9, 4] and in papers referred to there.

4 Filtered Related Work. Focusing on Trends

The sieve is double: a) vertical (problem-specific: e.g., learning based on right-brain tactics, heutagogy) and horizontal (domain-specific: e.g., continuum of care or cultural differences in e-Nursing). Here is reviewed only work regarding learning *per se*, principles of teaching (specificity of nursing included), and popular e-Learning paths (technologically advanced but within the dominant paradigm). To impair redundancy, only few references are retained from [35] and from [5] were related work was scrutinised less than a year ago.

Learning. Even in the rather deductive and apodictic cognitive environment of college-level mathematics, inductive reasoning is vital: "The primary goal [...] is to define the skill threshold necessary [...]. We have discovered two salient themes in the literature concerning what this means precisely. The first is the knowledge [...]. The second theme concerns the skills and abilities [...]. Abilities are attributes that affect the ability to perform a task, such as manual dexterity and inductive and deductive reasoning" [25].

The influence of affective processing in education was stressed recently in a dedicated workshop [16].

Conventional approaches model the learner profile for time spans between weeks and decades. For instance, in a recent substantial work [19] "Learners are assessed by several systems during their life-long learning. Those systems can maintain fragments of information about a learner derived from his learning performance and/or assessment in that particular system. Customization services would perform better if they would be able to exchange as many relevant fragments of information about the learner as possible".

A telling pointer that teaching is useless if the learner does not understand is found even in dogmatic settings, as shows the Augustinian-style of teaching: "Augustine provides two models - one for poor teaching and one for good teaching. Faustus was a poor teacher because he acted as an authority communicating "truth" externally. Ambrose became a good teacher because he pointed to the authority of truth discovered by learners within themselves" [29]. Indeed, Augustine, the first saint to dispose of a web page, was maieut (in *De magistro* he proposes dialogue as teaching practice) and forerunner of *learning through transforming experiences*: "This I remember; and have since observed how I learned to speak. It was not that my elders taught me words [...]; but I, longing by cries and broken accents and various motions of my limbs to express my thoughts, that so I might have my will, and yet unable to express all I willed [...] did myself" [39].

Non-Algorithmic Paths. The only relevant syntagm containing the attribute "non-algorithmic", defined in the FreeDictionary, is "non-algorithmic procedure": "A method of problem solving using exploration and trial and error methods. Heuristic program design provides a framework for solving the problem in contrast with a fixed set of rules (algorithmic) that cannot vary". No sign of a new paradigm.

Life-Long Learning and Andragogy. (The two terms are slightly different.) Rooted in ancient Greek education - maieutics is a blatant instance -, andragogy, as the "process of engaging adult learners in the structure of the

learning experience" (http://en.wikipedia.org/wiki/Andragogy) is "changing perceptions of adult learning theory and changing minds in academia" [38]. "Andragogy, initially defined as "the art and science of helping adults learn," has taken on a broader meaning since Knowles' first edition. The term currently defines an alternative to pedagogy and refers to learner-focused education for people of all ages. The andragogic model asserts that five issues be considered and addressed in formal learning. They include (1) letting learners know why something is important to learn, (2) showing learners how to direct themselves through information, and (3) relating the topic to the learners' experiences. In addition, (4) people will not learn until they are ready and motivated to learn. Often this (5) requires helping them overcome inhibitions, behaviors, and beliefs about learning" [17].

Another recent, very relevant comment in a specialised journal: "Learning is not restricted to the classroom and to formal learning inside learning institutions, it [...] happens throughout life, at work, play and home. In the modern knowledge-intensive era, life-long competence development has become a major challenge to our educational systems that have not changed their educational policies and pedagogical models to support life-long learning. There is an increasing demand for new approaches towards fostering life-long learning perspectives" [32].

An essential work for this paper is [15], where the following quotations come from. In this manual with real-world best practice examples: "E-learning will be considered any pedagogy (andragogy) that utilizes the Internet for communication". "E-learning should not be confused with distance education. Distance education is a program format in which the learners and instructors are geographically separate. While E-learning can be used in this format, it can also be used in an onsite program. Some programs allow learners the flexibility of moving in and out of f2f, distance education and E-learning through out their academic career". "Another form of E-learning is independent study". "The main theoretical bases upon which E-learning revolves are andragogy and constructivism. Andragogy is a term that refers to the teaching methodology that best facilitates learning in the adult. Constructivism refers to the belief that learning occurs as a result of the learner thinking about and interacting with the subject matter". "Adult learners tend to be self-directed in their learning and desire situations in which they can control their own education. The adult learner brings certain life experiences to the classroom that should be acknowledged as a frame of reference. They also require relevance in the content being studied. The information needs to be relevant for the adult to fully appreciate the need for the learning. These characteristics result in motivation for the adult learner to continue in their academic pursuits". "Constructivism focuses on the concept of knowledge construction versus knowledge transmission [...]. The basic focus of constructivism is that the learner interacts with the content being learned. This allows the learner to develop meaning about the content being learned within an environment that is/represents reality. In essence, the learner may acquire an understanding of basic principles and concepts by examining them within their natural environment" [15].

A recent collection of basic papers in this field is [34]. One of them [20] asserts: "We define remote in terms of geography, culture, language and telecommunications. One might think that the growing availability of 'open content' would make this an easy task, however our initial trials show this to be incorrect. Many of the current models for open content are not flexible enough to meet the demands of supporting 'meaningful learning'[...]. In order to investigate the individual's learning environment we undertook a very brief review of the current technology tools available on the desktop, that could assist learners in their tasks associated with personal knowledge management. There was nothing available that offered integrated support for knowledge management within a learner's personal domain that was available for use online and offline".

Heutagogy. Thus, a new concept emerged: *heutagogy*, "the principle of teaching based upon the concept of truly self-determined learning. It is suggested that heutagogy is appropriate to the needs of learners in the twenty-first century, particularly in the development of individual capability, individualised learning and independent learning using the internet-based systems including multimedia, virtual learning environments, online assessments and social software" (http://en.wikipedia.org/wiki/Heutagogy). "While Malcolm Knowles contributed greatly to our understanding of the limitations of pedagogy when it came to adult learning by defining andragogy, [...] andragogy did not go far enough. Any examination of learning experiences and curricula designed around andragogical principles certainly demonstrated the capacity for linking into the adult experience and recognised the advantages of self-directed learning. However, curricula were still very much teacher-centric with little opportunity for any real involvement [...] by the learner. [...] Action research allows experimentation with real world experience where learning is in the hands of the participants. [...] This is as close to real world learning as one can get in a controlled setting [...] doctoral students undertaking action research theses have progressed from pedagogical, then andragogical to heutagogical learning" [27].

Nursing in the "Continuum of Care". Tel-*eNursing* Practice is implemented in this continuum: "What was thought to be a quick and limited quality performance improvement (QPI) project between a few departments

providing telephone nursing services became an involved organization-wide QPI project to standardize telephone practice. Initially, telephone nursing practice, expanded roles of nurses, and extent of healthcare professionals using the telephone to provide care from afar were not well understood. Lack of standardization and quality and risk management issues were identified" [33].

From the stance of the approach proposed in the next section, as well as relevant for the main trends in this area and suggesting the signs of a paradigm shift is [44], excusing the long quotations that follow: "The Science of Nursing moved from a deterministic to a postmodern philosophical approach without a transitional period. Thus, to the expert observing the nursing profession two dogmas emerged, one related to healing of the body and the second to healing both mind and body". "Jean Watson in "Postmodern Nursing and Beyond" wrote: "The art and science of human caring and healing can be considered in some ways autopoietic; that is, it has been and is a discipline that is making itself"". "Leading the vehicle of the Science of Nursing into the 21st century, nurses need to escape from the barren professionalism, from the critical pathways, from the mechanistic nursing process, in order to discover ourselves through the daily practice, creating through nursing actions as exactly as the artist to create new shapes or expressions of Nursing Art even if the new creation(s) are characterized as skilfulness of communication, as moral interventions, as a teaching of self caring, nevertheless, they will include the Entirety in each part". "Socialization into nursing profession is the understanding of the world from inside, from our perception, thought, practice and actions interacting with metaholistic knowledge, and resisting the strict fact and deterministic model of the structure of the Human Being". "If anyone insisted on a deterministic approach to the changes described above, they would fail to adapt. The person using deterministic thinking pursues a vigorous analysis based on laws of foresight, of stability, and certainty, which result in a failure to adapt" [44].

Some of the ideas presented above are already carried out in practice, for instance by AACN (American Association of Critical Care Nurses) [26]: "When patient characteristics and nurse competencies match, patient outcomes are optimized, according to The Synergy Model [...] the most widely applicable framework for nursing practice".

New technology and old models. "Marshall McLuhan first noted the tendency to use new technologies in the model of the old. We have seen early examples (especially with remote classrooms) of teaching and learning that has hardly changed despite the investment of large sums of money and effort in new technology. Perhaps what is missing is new pedagogy that drives the development of new learning and assessment activities. In this topic we explore connectivism, heutagogy, e-Learning 2.0, and other ideas about formal and informal learning in the net-infused era" [2].

The "Google-Thermometer" shows that the indisputable superstar of e-Learning is the *learning object*, "a digital object that is used in order to achieve the desired learning outcomes or educational objectives" [41] organized in vast *repositories*, "LOR"s [18]. "Although the term, "learning object" originated from the notion of "object-oriented" computing and programming, which suggests that ideal way to build a computer program or anything digital is to assemble it from standardized, small, interchangeable chunks of code, the approach is somewhat different in an e-learning setting. In this case, learning management systems [...] could be considered large meta-objects, that contain spaces for the incorporation of granular objects [...]. Large repositories of learning objects are now available [...]. Although this tactic offers greater access and availability, they are not always easily navigated, nor is there a uniform system for classifying them" [41]. The same study concludes: "Change in the area of distance learning is rapid, and without a solid connection with underlying learning theories, the use of learning objects quickly becomes a function of the technology rather than the desired learning outcome. [...] As learning objects continue to evolve and new uses are found for them within online courses, it will be necessary to utilize a standard taxonomy/classification scheme. The need for this is urgent, because without it, learning object repositories will never be able to fully realize their mission of making learning objects available, easily retrievable, and sharable" [41].

Similar ideas about the problems of LORs: "the combination of prior knowledge analysis and a personal recommender system has a high potential to bridge the gap between the distributed resources and distributed self-directed learners who have the burden to choose suited learning activities and resources" [31]. "Globalization amounts to a massive downgrading of local context and offers the prospects of an unbounded, pervasive knowledge domain. Learning, however, if it is not restricted to professional training, consists in social processes, developing in multiple formats and channels of instruction and feed-back. Such events require a given location and a specific horizon of expectations. Learning objects may become the substitutes of text books, but this does not resolve the central challenge of education: mediating abstract knowledge and embodied, contingent patterns of expertise" [28].

It sounds familiar: no e-technology could solve the inherent problems of formal education, staying anchored in the old way of thinking. Worse, LORs inherited from their object-oriented model the technocentric stance

(focusing on "*reuse*") instead of the anthropocentric one (focusing on the very "*use*"). This unfortunate approach can be easily proved through a Google search: {"learning objects repositories" "adult learning"} returns about 91 results, whereas replacing "adult learning" with "andragogy", returns 9 results, and replacing it with "reuse", returns 1060 results. Hence, the old paradigm is more than ten times stronger!

5 A Two-Way Line of Attack

Before presenting the approach ("*Attack*"), the meta-approach ("*Two-Way Line*") should be explained.

Why a Two-Way Line? In brief, because a fresh concept in applied AI must fight at the same time to defend its intrinsic *scientific value* (convenience, purpose, transdisciplinary potential, propensity to follow trends, and so on) and its immediate *pragmatic relevance* (implementation effectiveness, user acceptance, downward compatibility, interoperability, etc.). To avoid both the vicious circle implied (relevance is based on implementations, but users are reluctant when previous relevance is lacking), and the vulnerability of non-validated theory, it is necessary to start also the other way around, exploiting the experimental models as "application surrogate". Consequently, after outlining the straight approach, a "reverse" one should be added to keep links to real-world problems, but solving yet toy ones. Toy problems are necessary because "it is generally accepted that academic research in East-European countries is still limited not by scientific potential but rather by financial or logistic boundaries" [12]. Thus, "a solution in search of a problem" is unaffordable and the caveat in [5] is valid: "a *VISON* (*VI*rtual *SO*cratic *N*urse) could be designed. However, its architectonics could be set up only if a genuine social command could be identified in the future." That is why, beyond the paper target, design-space dimensions are described and some mechanisms are briefly looked at.

Approach. Some macro-features are self-explaining: microcontinuity, transdisciplinarity, modularity, genericity, interoperability, idoneity and adhocracy ("change, from exception to rule" [3]). Corollary: the approach - or, at least, its language - is taken from generative programming (known also as meta-programming, mainly when it "leverages a computer language's power" [42]): "design space", "concerns" (here is used the term preferred in captology: "dimensions"), and "aspects".

To save space and to drop redundancy, the approach instantiated to e-Learning is adapted and abridged after [35]. Thus, the approach should be:

a) workable within the scope of a university research undertaking;

b) affordable for users with scarce resources (above all, individual "ageless learners");

c) suitable for setting up a test field for the architectonics of HA, PA and alike agent-oriented software;

d) stepwise (based on successive prototyping, with a pace depending on the progress of the projects involved).

Hence, as corollary of the meta-approach, there is no intention to develop a virtual Socrates outside the two projects mentioned. In other words, any possible *e-Socrates* should have an "e-DNA" akin to either HA, or PA, their crossover evolving in line with the results of both undertakings. Examples: neither the disk jockey nor the guitar teacher will clone themselves, but perhaps a virtual musicologist should do it if it could be more effective when its recently acquired knowledge - i.e., its dynamic ontology - is encrypted into its genotype; conversely, a HA experimental model would take over an enhanced sense of time - from its PA peer - if it improves its self-awareness. However, from the very beginning, both agents use the same "Simon-type" metrics to assess (agent or human) learning.

Since a maieutic process implies an "one to one" relationship, any learner profile is rather futile (a maieut interacts with a *person*, not with a *profile*) but the implicit learner meta-profile [35] is valid for generic agent architectures: a (highly motivated, very busy, pragmatic, goal-oriented, practised in Googling), adult. As well, knowledge (even static one) emerges as "coarse-grain", fuzzy, revisable, and highly personalised. Thus, a co-evolution of the pair is decidedly desirable: agent and human should learn together. In eNursing it is even more clear-cut: they are compelled to learn from each other: "the ideal phenomena of the classic deterministic consideration are becoming chaotic and complex phenomena and are frequently found in nursing practice. Under the postmodern consideration, in order to approach the complex phenomena of nursing practice, nurses must follow the example of Plato's prisoner and free themselves to emerge into the real world. Today, there are no correct answers, but simply there are better approaches; today there are no rules, there are simply ways" [44]. Indeed, neither teacher nor learner can be confined to algorithmic methods.

In short, the maieutic learning process should be fundamentally non-algorithmic, mostly stimulus-driven, and involving, as much as possible, right hemisphere tactics (affective computing included).

Though, "non-algorithmic" does not imply "sub-symbolic", because:

a) Symbolic processing is unavoidable in learning [1]).

b) Anthropocentric interfaces require symbols.

c) Massive parallelism is hardly affordable with scarce resources [5].

d) Symbols are implied by "Piaget's distinction between assimilation and accommodation as mechanisms of learning and development" [1].

6 Specific Design-Space Dimensions and Mechanisms

The design space for a maieutic agent is a (more or less extended) subset of the Cartesian product:

$$\mathbf{S}\text{Socrates} = \mathbf{S}\text{Maieutics} \times \mathbf{S}\text{Agents}$$

SAgents. The design space is set up depending on the agent type. Since any kind of education involves *persuasion* (teachers - and coaches even more - to be effective must be *convincing*), a suitable framework is the design space for persuasive agents, introduced and supplemented in [3] for the emotivity dimension of pathematic agents, and instantiated in [9] for emotional Disneyland characters (dedicated to paediatric aims):

$$\mathbf{S}\text{PersuasiveAgent} = \mathbf{S}\text{Captology} \times \mathbf{S}\text{MedicalInformatics} \times \mathbf{S}\text{Agents}$$

where **S**Agents is meant to highlight the idea that agent-oriented applications are not *products* but a*ctive knowledge* [3]. (This design space was the framework for all virtual therapists mentioned in Section 3.)

SMaieutics. The cardinal dimensions for e-Maieutics that can be inferred from its rationale are:

$$\mathbf{S}\text{Maieutics} = \{life-long\ learning,\ bounded\ rationality,\ affective\ computing,\ ethical\ behaviour\}$$

The first two dimensions are not yet implemented as such in any experimental model.

- *Life-Long Learning*. At this stage, only some basic aspects can be considered as candidates for initial conceptualising:

  a) The "Continuum of Care" assigns a very clear-cut - and sometimes dramatic - meaning to "life-long".

  b) While *learner profile* is rather futile, *learning history* is vital (mainly, but not only, as feedback for the teacher).

  c) Andragogy is a must, but the obvious trend is towards heutagogy.

  d) This involves user-driven dialogue: strategic agent proactiveness is replaced stepwise by a larger pallet of services.

  e) General features of learning as cognitive process (e.g., fundamentally non-algorithmic, mostly stimulus-driven, motivation dependent) are substantially amplified in maieutic context.

- *Bounded Rationality*. Since it is a rich and versatile concept [40, 30], often applied as principle, it is awkward to be used as design space dimension, above all in e-Nursing: "The problem is neither to admit that for any medical act (and for even stronger reasons as regards nursing) "just in time" is a sine qua non condition, nor that bounded rationality is the only practical means to achieve it [24]. Nevertheless, there is a double hindrance, due to a yet prevalent mentality: a) therapeutic decision-making is an exclusively human attribute; b) non-algorithmic software is - if not nonsensical - applicable at most to toy problems" [5]. As a result, bounded rationality will not be just a design space dimension but a fascicle of interrelated, versatile, and highly application-dependent features. Some examples: learning strategies (are autodidacts welcomed?), negotiation strategies (dialectics and the art of compromise), any feature meant to reduce complexity (architectural, cognitive, and structural: "zero-overhead rule"). Any bounded rationality dimension must have aspects able to cover a wide range of values and to be personalised any time [33]. Though, to preserve the core of bounded rationality, all dimensions or aspects should manifest coarse granularity.

- *Affective computing*. Defined in Wikipedia as branch of AI "that deals with the design of systems and devices which can recognize, interpret, and process emotions" the term could mislead and is employed here instead of its partial synonym "captology" only because it is about ten times more widespread - both terms, coined in the 90's, being yet in a fuzzy conceptual relationship. This dimension is needed for theoretical and practical reasons, namely to catalyse the emergence of self-awareness in HA or of a "thick time" sense of duration [8] in PA, agents should manifest stepwise human-like behaviour, affective features included. Here micro-continuity can help because "not the *anthropomorphic feature* itself has to be replicated , but its *appearance* - firstly forged, later more genuine" [11]. A basic such feature is *emotivity*: to be *convincing* (see above), teachers must be *credible*; that involves manifesting emotivity (as humans) or imitating it (as agents). Hence, when "designing requirements for the persuasion-related features [...], (e.g. credibility, rhetoric, agent voice pitch, rise/fall of [...] agent emotions) one has to separate out the relevant concerns" [5].

- *Ethics*. Given that the ethical facet of any educational or medical act is crucial, the need to reflect it in practice became almost an axiom. Moreover, the teacher-learner relationship is special: long lasting, multifaceted, and difficult. In eNursing the need is even more urgent because nurse-patient relationship involve also other persons (family, physicians, etc.). Perhaps the main reasons why ethics is vital for eNursing effectiveness is that nursing is based on *persuasion* (see above), and on the "*Primum non nocere*" principle. Ethical issues in captological applications were addressed in [3] and recently in [10], proposing an ethical potentiometer that "can indicate what we have called "variable ethical rigor", i.e. enabling us to switch on a scale from one ethical level to another. The ethically divisive topic we have chosen is the clandestine practice of sending a patient subliminal messages". "The different elements of ethics required in the design process need to correspond to categories of ethics as system, expressing various degrees of rigor. At one extremity, [...] strict deontological form of ethics (total intransigence: standards can never be broken, regardless of causing pain), at the other [...] Epicurean act-based pragmatism (the ethics that talks in terms of "pros and cons "); and somewhere in between one can place rule-based utilitarianism (rules are set in place only if always following them proves to be beneficial)" [10]. In [5] the restrictions from [10] were adapted for virtual nurses.

Mechanisms [8]. The software toolbox *AGORITHM* (*AG*ent-*OR*iented *I*nteractive *T*ime-sensitive *H*euristic *M*echanisms) includes explicitly not object-oriented mechanisms: antientropic self-cloning; decision-making with future contingents; exception-driven reactivity; clone-based "thick time"; dynamic mini-ontology; visual mini-ontology. The first three mechanisms are implemented (verified as operational in at least one experimental model for agent-based e-Learning applications), the fourth is in early development, and the last two are still in a syncretic stage of conceptualizing. All are based on API functions callable from customary development environments and on the first FIPA standard [21] defining the agent as *process*.

7 Conclusions and Future Work

Without validation the assertions below, albeit most of them factual, should be seen as estimations:

- *e-Maieutics*, as innovative concept is endorsed by the *context* (uncertain, rapidly changing environments), *targets* (personalised dynamic knowledge), *nature* (non-algorithmic information processing), and *methods* (andragogy tending towards heutagogy), of learning in the era of "computing as interaction".

- The proposed two-way (meta-)approach is:

  a) workable within the scope, and affordable with the scarce resources of university research undertakings;

  b) suitable for setting up a test field for the architectonics of protensional agents, self-referencing agents and alike non-algorithmic software.

- The design space outlined for Socratic agents opens promising paths for both e-Learning and agent-oriented software; the dimensions and mechanisms proposed prove to be a good test-bench for related research.

Future work:

a) As regards virtual maieuts, until a social command for e-Maieutics is in sight, future work is confined to developing/improving mechanisms (for instance, now is modelled a mechanism for evaluating "Simon-type learning", based on a simple time derivative of task completion duration).

b) For the related projects, the medium-range targets are: "thick time" for protensional agents and a mirror-test mechanism for self-referencing agents. Long-range targets are set up in the PhD thesis in preparation [22].

Since the conclusions are far from being apodictic, while future work is tentative, they could be cut *in nuce* along the landmarks of the well-known BDI agent architecture. *Beliefs*: the paradigmatic shift becomes urgent. *Desires*: the paper should generate a debate to ease such a shift. *Intentions*: keep walking (both journey and destination are worth the struggle).

Acknowledgments. Emil M. Popa and Ioana Moisil sustained in most conventional way some most unconventional thought; the truth of the next sentence is due entirely to their intellectual tolerance: This work was supported by the Ministry of Education and Research through Grant CNCSIS 33/2007.

References

[1] J. R. Anderson, L. M. Reder, H. A. Simon, *Applications and Misapplications of Cognitive Psychology to Mathematics Education*. Texas Educational Review, 2000.

[2] T. Anderson, MDDE 663: Emerging Issues In Distance Education Technologies. *Course Offerings and Official Syllabi*, Athabasca University, Centre for Distance Education, Edmonton, AB Canada, Fall/2006.

[3] B. E. Bărbat, *Agent-Oriented Intelligent Systems*. Romanian Academy Publishing House, Bucharest, 2002 (in Romanian, "Grigore Moisil" Prize of the Romanian Academy).

[4] B. E. Bărbat, The Impact of Broad-Band Communication upon HMI Language(s). (Chapter 7.) Communicating in the world of humans and ICTs. (Chapter 8.) in COST *Action 269. e-Citizens in the Arena of Social and Political Communication* (L. Fortunati, Ed.), 113- 142, EUR21803, Public. of the European Comm., Luxembourg, 2005.

[5] B. E. Bărbat, From e-Learning to e-Nursing, Applying Non-Algorithmic Paths. *Medinf 2007 Workshop: E-Learning aspects in medicine and nursing*. 29th International Conference of the Romanian Medical Informatics Society, Sibiu, 2007. (To appear on CD).

[6] B. E. Bărbat, Natural time for artificial agents. *Proc. of Exploratory Workshop on NL-Computation*, Baile Felix, May 15-17, 2008 (Invited paper.)

[7] B. E. Bărbat, R. D. Fabian, C. Brumar, R. S. Muntean. Bounded Rationality and Approximation in Modern Artificial Intelligence. *Proc. of the International Workshop New Approaches, Algorithms and Advanced Comp. Techniques in Approximation Theory and its Applications* (D. Simian, Ed.), 2007. (In print.)

[8] B. E. Bărbat, A. V. Georgescu, E. M. Popa. Time for Bodiless Agents. A Software Engineering Approach to Self-Awareness. (Submitted to *SIWN Int. Conf. on Self-organization and Self-management in Computing and Communications*, 2008.)

[9] B. E. Bărbat, D. Luca. Emotional Disneyland Characters for Paediatric Purposes. *MIE2003 Medical Informatics Europe. The new navigators: from professionals to patients*, CD St Malo, 2003.

[10] B. E. Bărbat, A. Moiceanu, H. G.B. Anghelescu. Enabling Humans to Control the Ethical Behaviour of Persuasive Agents. Chapter 14 in E. Mante-Meijer, L. Haddon. E. Loos (Eds.) *The Social Dynamics of Information and Communication Technology*. Ashgate, Aldershot, UK. (To appear: August, 2008.)

[11] B. E. Bărbat, A. Moiceanu, I. Pah. Gödelian Self-Reference in Agent-Oriented Software. *Proc. of the 11th WSEAS Int. Conf. on COMPUTERS* (ICCOMP '07) (N.E. Mastorakis et al, Eds.), 92-97, Agios Nikolaos, Crete, 2007.

[12] B. E. Bărbat, A. Moiceanu, S. Plesca, S. C. Negulescu. Affordability and Paradigms in Agent-Based Systems. *Computer Science Journal of Moldova*, 15, 2(44), 178-195, 2007.

[13] B. E. Bărbat, S. C. Negulescu, A. E. Lascu, E. M. Popa. Computer-Aided Semiosis. Threads, Trends, Threats. *Proc. of the 11th WSEAS Int. Conf. on COMPUTERS (ICCOMP '07)* (N.E. Mastorakis et al, Eds.), 269-274, Agios Nikolaos, Crete, 2007.

[14] B. E. Bărbat, S. C. Negulescu, S. Plesca. Emergence as Leverage and Non-Algorithmic Approaches in Agent-Oriented Software. *Studies in Informatics and Control Journal*, 16, 4, 321-332, 2007.

[15] T. J. Bristol, *Evidence-based E-Learning for Nurse Educators*. Center for Health Workforce Planning Bureau of Health Care Access Iowa Department of Public Health, 2006

[16] P. Chalfoun, S. Chaffar, C. Frasson. Predicting the Emotional Reaction of the Learner with a Machine Learning Technique. *Workshop on Motivational and Affective Issues in ITS*. (G. Rebolledo-Mendez, E. Martinez-Miron, Eds). 8th International Conference on ITS, 13-20, 2006.

[17] M. L. Conner, Andragogy and Pedagogy. *Ageless Learner*, 1997-2004 http://agelesslearner.com/ intros/ andragogy.html

[18] G. Doctor, S. Ramachandran. Learning Object Repositories An Emerging Knowledge Management Tool for Sharing and Reusing Learning Resources. *ICFAI Journal of Knowledge Management*, V, 2, Icfai University Press, Hyderabad, 2007 (http://hdl.handle.net/123456789/294).

[19] P. Dolog, M. Schäfer. A Framework for Browsing, Manipulating and Maintaining Interoperable Learner Profiles. In *User Modeling* (J.G. Carbonell, J. Siekmann, Eds.), Springer Berlin / Heidelberg, 2005.

[20] C. Esslemont, 'Bridging the abyss': open content to meaningful learning. *OpenLearn Researching open content in education Proceedings of the OpenLearn2007 Conference* 30-31 October 2007 Milton Keynes, UK Patrick McAndrew and Jo Watts (Editors)

[21] FIPA TC Agent Management. FIPA *Agent Management Specification*. Standard SC00023K (2004/18/03). http://www.fipa.org/specs/fipa00023/ SC00023K.pdf

[22] A. V. Georgescu, *Agent-Oriented Semiosis for Protensity Messages*. Application in Musicology. (PhD Thesis in preparation.)

[23] A. V. Georgescu, E. M. Popa, B. E. Bărbat. Time-Aware Agent as Virtual Guitar Teacher. (Submitted to *SIWN Intern. Conf. on Self-organization and Self-management in Computing and Communications*, 2008.)

[24] G. Gigerenzer, A. Edwards. Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, 327, 741-744, 2003.

[25] P. Golfin, et al. *Strengthening Mathematics Skills at the Postsecondary Level: Literature Review and Analysis*. U.S. Department of Education. Office of Vocational and Adult Education. Division of Adult Education and Literacy, 2005.

[26] S. R. Hardin, R. Kaplow. *Synergy for Clinical Excellence: The AACN Synergy Model for Patient Care*. Jones and Bartlett, Sudbury, Mass., 2005.

[27] S. Hase, C. Kenyon. Heutagogy: A Child of Complexity Theory. *Complicity: An International Journal of Complexity and Education*, 4, 1, 111-118, 2007.

[28] H. Hrachovec, e-Learning Nudism: Stripping Context from Content. In *Mobile Understanding: The Epistemology of Ubiquitous Communication* (K. Nyíri, Ed.) Passagen Verlag, Vienna, 103-110, 2006.

[29] R. M. Jacobs, O. S. A. Augustine's Pedagogy of Intellectual Liberation: Turning Students from the "Truth of Authority" to the "Authority of Truth" in K. Paffenroth and K. L. Hughes, *Augustine and Liberal Education*. Aldershot, England: Ashgate, 117, 2000.

[30] D. Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*. Lecture (when receiving Nobel Prize; revised version). Stockholm, Nobel Foundation, 2002.

[31] M. Kalz, et al. Wayfinding Services for Open Educational Practices. *International Journal of Emerging Technologies in Learning*, 3, 2, 2008. (http://hdl.handle.net/1820/1155.)

[32] R. Klamma, et al. Social Software for Life-long Learning. *Educational Technology and Society*, 10 (3), 72-83, 2007.

[33] M. L. Larson-Dahn, Tel-*eNursing* Practice: Setting Standards for Practice Across the Continuum of Care. *Journal of Nursing Administration*, 32; Part 10, 524-530 (direct.bl.uk/research/2E/21/ RN120920856.html, 2002.

[34] P. McAndrew, J. Watts (Editors) OpenLearn: Researching open content in education. *Proceedings of the OpenLearn2007 Conference*, Milton Keynes, UK, 2007.

[35] I. Pah, A. Moiceanu, I. Moisil, B. E. Bărbat. Self-Referencing Agents for Inductive Non-Algorithmic e-Learning. *Proc. of the 11th WSEAS International Conference on COMPUTERS (ICCOMP '07)* (N.E. Mastorakis et al, Eds.), 86-91, Agios Nikolaos, Crete, 2007.

[36] E. M. Popa, A. V. Georgescu, B. E. Bărbat. E-Learning with Protensional Agents: Playing Guitar. (Submitted to *12th WSEAS International Conference on COMPUTERS (ICCOMP '07)*, Heraklion, Crete, 2007.)

[37] L. B. Resnick, *Education and Learning to Think*. National Academy Press, Washington, DC, 1987.

[38] J. D. Royer, The New Distance Learning: Changing Perceptions of Adult Learning Theory and Changing Minds in Academia. *Journal of Business and Public Policy* 1, 4, 1-15, 2007.

[39] Saint Augustine. *The Confessions of St. Augustine*. The Harvard Classics. 1909-14. The First Book

[40] H. A. Simon, *Models of Bounded Rationality*. MIT Press, Cambridge, 1997.

[41] S. Smith Nash, Learning Objects, Learning Object Repositories, and Learning Theory: Preliminary Best Practices for Online Courses. *Interdisciplinary Journal of Knowledge and Learning Objects* (A. Koohang, Ed.) Vol. 1, 2005.

[42] D. Spinellis, Rational Metaprogramming. *IEEE Software*, 25, 1, 78-79, Jan/Feb, 2008.

[43] N. Triola, Dialogue and Discourse: Are We Having the Right Conversations? *Crit Care Nurse*, 26, 1, 60-66, 2006.

[44] S. L. Van Sell, I. A. Kalofissudis. *The evolving essence of the science of nursing. A complexity integration nursing theory*. E-book online, ISSN 1108-7366, ICUS and Nursing Web Journal, 2002.

Boldur E. Bărbat
"Lucian Blaga" University of Sibiu
Faculty of Sciences
Ion Raţiu St. 5-7, 550012, Sibiu, Romania
E-mail: bbarbat@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 55-66

On the Representation and the Stability Study of Large Scale Systems

Plenary invited paper & workshop invited key lecture

Pierre Borne, Mohamed Benrejeb

Abstract: Aggregation techniques, associated to Lyapunov functions or to vector norms, enable to conclude to large scale systems stability if their less or equal order comparison systems are stable. Linear conjecture is then satisfied if the comparison system is identical to the system itself.
Practical Borne-Gentina stability criterion, applied to continuous systems, generalizes the use of Kotelyanski lemma for non linear systems and defines large classes of processes for which linear Aizerman conjecture is satisfied.
The stability approach is applied for multimodel control systems and for TSK fuzzy models when Benrejeb arrow form characteristic matrix is used.
Keywords: continuous large scale systems, stability, Borne-Gentina criterion, multimodel control system, TSK fuzzy model, Benrejeb arrow form matrix.

1 On the choice of Lyapunov function and vector norm for stability study. Basic idea

The stability theory aims at drawing conclusions about the behaviour of a system without actually computing its solution trajectories; Lagrange (1788) concluded that, in the absence of external forces, an equilibrium state of a conservative mechanical system is stable provided that it corresponds to a minimum of the potential energy. Then, for a long period, stability studies have been limited to conservative mechanical systems described by Lagrangian equations of motion.
The quantum advance in stability theory that allowed one the analysis of arbitrary differential equations is due to A. M. Lyapunov (1892), who introduced the basic idea and the definitions of stability that are in use today and proved many of the existing fundamental theorems [31]. The concept of Lyapunov stability plays an important role in control and system theory.
Solutions of stability problems risen from large scale systems, have been approached [9, 21, 25, 35] by singular perturbation method to achieve dimensionality reduction by considering separately slow and fast system, Bellman's concept (1962) of vector Lyapunov functions and Bailey and Siljak approaches to find conditions and interactions under which stability property of the overall system is inferred from stability properties of subsystems, and Robert's concept (1964) of vectors norms, and its multilevel application by Borne et al (1973) and Borne and Gentina (1974). Lasalle and Lefschetz (1961) developed non-Lyapunov stability investigations, generalized by Weiss and Infante (1965). Then many results on conditions for different types of practical stability have been given by Weiss (1967), Weiss and Infante (1965,1967), Michel and Porter (1972), then Grujic (since 1970).
Linear system stability study generally leads to necessary and sufficient conditions and doesn't depend on the system representation. The task is different for non linear systems with or without uncertainties, for which only sufficient conditions are given, then their stability domains depend on the choice of both description of the studied system and the used stability method [3, 4, 5, 6, 34].
To lead stability study, one can start with a system description (resp. stability methods), then choose an adapted analysis method (resp. system description). Then, the problem of Lyapunov functions construction and the problem of determining the largest stability domain, for example, can be transformed by a suitable system description choice [5]. Nyquist frequencial diagram, characteristic matrix or its characteristic polynomial, for example, are useful to apply respectively Nyquist criterion, Kotelyanski lemma, or Routh criterion.
The satisfaction of Aizerman conjecture for a non linear system sometimes needs [3, 20] only a change of the state vector transforming the characteristic matrix when sufficient stability conditions are in test.
For large scale systems, generally described in state space, stability conditions are obtained, for the whole system or for subsystems and interactions [17, 18, 19, 21, 24, 36]. Several approaches are considered in the literature using decomposition of a large scale system into subsystems, or of the vector Lyapunov function or the vector norm taking into account specific theoretical or physical properties of the process. In such a case, the system can be considered, for example, as an interconnection of controllable, observable, or stable subsystems. Then the analysis and the synthesis problems can arise to the level to study the properties transfer from the subsystems (resp.

whole systems) to the whole system (resp. subsystems) [3].
In this paper, the influence of the state vector description and of the matrix characteristic, on the determination of the stability domain, is studied by Borne-Gentina stability criteria [9, 17, 18, 22, 23], based on the use of vector norms and of overvaluing systems of nonlinear large scale systems, for which characteristic matrices are in Compagnon, Frobenius, diagonal, Jordan [16], or Benrejeb form [11] and for vector norms leading or not to the satisfaction of Aizerman conjecture [20] to a comparison system.
The use of this basic idea is generalized to the stability study of multimodel system control [13, 26, 29, 34] and TSK fuzzy systems [1, 2, 4, 7, 12, 27, 28, 39, 40, 41, 42, 43, 44].

2 Stability conditions using the overvaluing continuous systems

Let us consider the vector differential equation:

$$\dot{x} = f(x(t),t),\ t > 0,\ x(t) \in R^n,\ f : R^n \times R+ \rightarrow R^n \tag{1}$$

It is further assumed that this equation has a unique solution corresponding to each initial condition. This is the case if f satisfies a global Lipschitz condition.
Let $s(t,t_0,x_0)$ denote the solution corresponding to the initial condition $x(t_0) = x_0$; s satisfies: $\dot{s}(t,t_0,x_0) = f(s(t,t_0,x_0),t)\ ,\ \forall t > t_0,\ s(t_0,t_0,x_0) = x_0$.
$x_0 \in R^n$ is an equilibrium state of the studied system if: $f(x_0,t) = 0, \forall t > 0$, then: $s(t,t_0,x_0) = x_0, \forall t \geq t_0 > 0$.
In other words, if the system starts at an equilibrium position, it stays there [3, 9].
If the equilibrium state under study is not the origin, one can redefine the coordinates on R^n in such a way that the considered equilibrium becomes the new origin.

Definition 1. The equilibrium state O is stable if, for each $\varepsilon > 0$ and each $t_0 \in R+$, there exist a δ, $\delta = \delta(\varepsilon,t_0)$ such that: $\|x_0\| < \delta(\varepsilon,t_0) \Rightarrow \|s(t,t_0,x_0)\| < \varepsilon, \forall t > t_0$.

Definition 2. The equilibrium state O is attractive if, for each $t_0 \in R+$, there is an $\eta(t_0) > 0$ such that: $\|x_0\| < \eta(t_0) \Rightarrow \|s(t,t_0,x_0)\| \rightarrow 0$ as $t \rightarrow +\infty$.

Definition 3. The equilibrium state is asymptotically stable if it is stable and attractive.

Definition 4. The equilibrium state is exponentially stable if there exit constants $r,a,b > 0$, such that: $\|s(t,t_0,x_0)\| \leq a\|x_0\| exp(-b(t-t_0)), \forall t > t_0, \forall x \in Br, Br = \{x \in R^n : \|x\| \leq r\}$.

Lyapunov stability
The concept of Lyapunov stability plays an important role in control and system theory [24]. If a system is initially in an equilibrium state, it remains in the same state there after. Lyapunov stability is concerned with the behaviour of the trajectories of a system when its initial state is near an equilibrium state.
Let $\dot{x} = f(x,t)$ continuous, $\forall t \in Z$ and $\forall x \in D \subset R^n$, and $v(x)$ a Lyapunov candidate function satisfying the Lyapunov conditions. If: $\dot{v}(x,t) \leq -\Psi(\|x\|)$, with Ψ continuous, definite and positive $\forall \|x\| \in [0,H]$, then the solution $x(t) \equiv 0$ is uniformly asymptotically stable.

Definition of the overvaluing continuous systems
Let us consider the class of the studied system described by the differential equation:

$$\dot{x}(t) = A\,(x(t),t)\,x(t) \tag{2}$$

where $x(t)$ is an n vector of the state space E defining the system at the time t, $A(x(t),t)$ represents an $n \times n$ matrix, elements of which are functions of both x and t.

Definition 5. The matrix $M_c(x,t)$ is a pseudo-overvaluing matrix of matrix A with respect to the vector norm p, if the inequality: $\dot{p}(x) \leq M_c\,(A\,(x(t),t))\,p(x),\ \forall x \in E$ and $t > 0$, is verified for each corresponding component. Consequently, the stability of the comparison system: $\dot{z} = M_c(x,t)z$, with the initial conditions such as $z_0 = p(x_0)$, implies the same property for the system: $\dot{x} = A\,(x,t)\,x$.

The determination of a pseudo-overvaluing matrix of a given matrix is presented in the appendix.
When a pseudo-overvaluing matrix M_c of a matrix A is defined with respect to a regular vector norms p, we can bring the following properties.

(i) The off-diagonal elements of matrix $M_c(A)$ are non negative.

(ii) If we denote by $R_e(\lambda_M)$ the real part of the eigenvalue of maximum real part of $M_c(A)$, it comes the inequalities: $R_e\left(\lambda_{A(x,t)}\right) \leq R_e(\lambda_M)$.

(iii) When all the real parts of the eigenvalues of $M_c(A)$ are negative, this matrix is the opposite of an M-matrix, i.e., it admits an inverse whose elements are all non positive.

When this inverse is an irreducible matrix, then $M_c(A)$ admits an eigenvector $u(x,t)$ relative to the eigenvalues λ_M and whose components are strictly positive.
In particular cases, the use of specific Lyapunov function deduced from vector norms [9, 17, 18, 19, 20, 21, 22, 23, 24, 25, 32, 35, 36] for the stability study of the initial system to the one of a comparison system is reduced for which the linear conjecture is satisfied.

3 Borne-Gentina stability criterion

Let us consider the linear process described in state space by: $\dot{x} = Ax$, where A is an $n \times n$ matrix. Hurwitz conditions applied to characteristic polynomial parameters of A, $A = \left\{a_{i,j}\right\}$, leads to the asymptotic stability domain in parameters space $a_{i,j}$. Kotelyanski elaborates a particular lemma, well adapted for stability study when off diagonal elements of matrix A are non negative.

Lemma 6. ***Kotelyanski lemma*** [17]. *The real parts of the eigenvalues of matrix A, with non negative off diagonal elements, are less than a real number μ if and only if all those of matrix M, $M = \mu I_n - A$, are positive, with I_n the n identity matrix.*
When successive principal minors of matrix $(-A)$ are positive, Kotelyanski lemma permits to conclude on stability property of the system characterized by A.

Borne-Gentina practical stability criterion [17]
Let consider the nonlinear continuous process described in state space by: $\dot{x} = A(.)x$; $A(.)$ is an $n \times n$ matrix, $A(.) = \left\{a_{i,j}\right\}$. If the overvaluing matrix $M(A(.))$ has its non constant elements isolated in only one row, the verification of the Kotelyanski condition enables to conclude to the stability of the initial system.

As an example, if the non constant elements are isolated in only one row of $A(.)$, Kotelyanski lemma applied to the overvaluing matrix obtained by the use of the n regular vector norm $p(x)$ with $x = [x_1, x_2, \ldots, x_n]^T$, such that: $p(x) = [|x_1|, |x_2|, \ldots, |x_n|]^T$, leads to the following stability conditions of initial system:

$$a_{1,1} < 0, \begin{vmatrix} a_{1,1} & |a_{1,2}| \\ |a_{2,1}| & a_{2,2} \end{vmatrix} > 0, \ldots, (-1)^n \begin{vmatrix} a_{1,1} & |a_{1,2}| & \cdots & |a_{1,n}| \\ |a_{2,1}| & a_{2,2} & \cdots & |a_{2,n}| \\ \vdots & \vdots & & \vdots \\ |a_{n,1}(.)| & |a_{n,2}(.)| & \cdots & a_{n,n}(.) \end{vmatrix} > 0 \tag{3}$$

This criterion is useful for the stability study of complex and large scale systems, such that the necessary condition of its application is satisfied or if the system parameters identification is imprecise.
The Borne-Gentina practical criterion applied to continuous systems generalizes the Kotelyanski lemma for non linear systems and defines large classes of systems for which the linear conjecture can be applied, either for the initial system or for its comparison system [3, 17, 20].

4 On the choice of characteristic matrix form

The determination of the largest complex system stability domain, based on the use of Lyapunov method associated with aggregation techniques using vector norms, depends on the choice of the description of the studied system which can be conditioned by mathematical or physical considerations.
Compagnon or Frobenius form matrix when its characteristic polynomial is known, and diagonal and/or Jordan form matrix when modes can be easily computed, are the main matrix canonical forms [16].

Let us consider the non linear continuous process whose model is in the controllable form, that matrices $A_i(.)$ in Compagnon form, of equation (2), are written as:

$$A_i(.) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & & 0 \\ 0 & & \cdots & 0 & 1 \\ -a_{i,0}(.) & & \cdots & & -a_{i,n-1}(.) \end{bmatrix} \tag{4}$$

$a_{i,j}(.)$ is a coefficient of the instantaneous characteristic polynomial $P_{A_i}(.,\lambda)$ of the matrix $A_i(.)$, such that:

$$P_{A_i}(.,\lambda) = \lambda^n + \sum_{l=0}^{n-1} a_{i,l}(.)\,\lambda^l \tag{5}$$

A change of base under the form [3]:

$$T = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_1^{n-2} & \alpha_2^{n-2} & \cdots & \alpha_{n-1}^{n-2} & 0 \\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_{n-1}^{n-1} & 1 \end{bmatrix} \tag{6}$$

allows the new state matrix, denoted by $F_i(.)$, to be in arrow form [3]:

$$F_i(.) = T^{-1} A_i(.)\, T = \begin{bmatrix} \alpha_1 & & & \beta_1 \\ & \ddots & & \vdots \\ & & \alpha_{n-1} & \beta_{n-1} \\ \gamma_{i,1}(.) & \cdots & \gamma_{i,n-1}(.) & \gamma_{i,n}(.) \end{bmatrix} \tag{7}$$

with:

$$\beta_j = \prod_{\substack{k=1 \\ k\neq j}}^{n-1} (\,\alpha_j - \alpha_k\,)^{-1}, \forall j = 1,2,...,n-1 \tag{8}$$

$$\gamma_{i,j}(.) = -P_{A_i}(.,\alpha_j), \forall j = 1,2,...,n-1 \tag{9}$$

$$\gamma_{i,n}(.) = -a_{i,n-1}(.) - \sum_{i=1}^{n-1} \alpha_i \tag{10}$$

and α_j, $j = 1,2,...,n-1$, are distinct arbitrary constant parameters.

This matrix in arrow form is called Benrejeb matrix [11]. In [13], three general kinds of those matrices are introduced: thin (A_{tAF}), generalized thin (A_{GtAF}) and thick A_{TAF} matrices considered in arrow form.

$$A_{tAF} = \begin{bmatrix} f_{1,1} & & f_{1,n} \\ & \ddots & \vdots \\ f_{n,1} & \cdots & f_{n,n} \end{bmatrix} \tag{11}$$

$$A_{GtAF} = \begin{bmatrix} f_{1,1} & & & f_{1,r} & \cdots & f_{1,n} \\ & \ddots & & \vdots & & \vdots \\ & & f_{r-1,r-1} & f_{r-1,r} & \cdots & f_{r-1,n} \\ f_{r,1} & \cdots & f_{r,r-1} & f_{r,r} & \cdots & f_{r,n} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ f_{n,1} & \cdots & f_{n,r-1} & f_{n,r} & \cdots & f_{n,n} \end{bmatrix} \tag{12}$$

$$A_{TAF} = \begin{bmatrix} F_{1,1} & & F_{1,r} \\ & \ddots & \vdots \\ F_{r,1} & \cdots & F_{r,r} \end{bmatrix} \tag{13}$$

where $F_{i,i}$ is an $n_i \times n_i$ matrix such that : $\sum_{i=1}^{r} n_i = n$, and $F_{i,j}$ an $n_i \times n_j$ matrix.
In [5], many physical linear and non linear systems are directly described by characteristic matrix in Benrejeb forms corresponding with a theoretical two hierarchical levels structure.
The determinant computation is as easy for the matrix A_{tAF} [3]:

$$|A_{tAF}| = (f_{n,n} - \sum_{i=1}^{n-1} \frac{f_{n,i} f_{i,n}}{f_{i,i}}) \prod_{i=1}^{n-1} f_{i,i} \tag{14}$$

as for the matrix A_{TAF}:

$$|A_{TAF}| = \left| F_{r,r} - \sum_{i=1}^{r-1} F_{r,i} F_{i,i}^{-1} F_{i,r} \right| \prod_{i=1}^{r-1} |F_{i,i}| \tag{15}$$

Borne-Gentina criterion, based on determinants computation, is then well adapted for systems described by characteristic matrices in Benrejeb form, A_{tAF}, A_{GtAF} and A_{TAF}, such that non linear elements are isolated in its last rows.

5 Application to multimodel control systems stability study

In this section, a result on the stability of non linear multimodel systems using pole assignment control, is presented [13, 15, 26, 29, 30, 34]. The controllable form is adopted for the state space representation for each model in the model basis. This representation allows finding a state feedback which, for the closed loop system, determines the same free state matrix for the whole set of models in the basis. The control strategy for the global system can then be chosen under the state feedback with parameters fixed by using a fusion of the different feedback parameters derived for each model. The fusion is carried out by taking into account the validity of each model.
In the non linear case, the free state matrix of the model is considered to be in the Compagnon form or, when this is not the case, its state matrix is transformed to the Benrejeb matrix in arrow form. This particular form allows having the non-constant elements of the free state matrix isolated in the last row, which makes it possible to establish a stability criterion for the non linear system in the multimodel approach.
Let us consider the process whose model is in controllable form:

$$\dot{x} = A_i(.)x + Bu \tag{16}$$

$$A_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & & 0 \\ 0 & & \cdots & 0 & 1 \\ -a_{i,0}(.) & & \cdots & & -a_{i,n-1}(.) \end{bmatrix}, B = \begin{bmatrix} 0 \\ \vdots \\ \\ 0 \\ 1 \end{bmatrix} \tag{17}$$

where $a_{i,j}$, $j = 1,...,n-1$, are bounded for finite values of their arguments, and a set of models M_i $i = 1,...,r$, in the model basis. The same pole assignment control law is carried out for each model M_i, and the state feedback parameters are denoted by L_i, In this case, the following result holds:

Theorem 7. *Let all n poles $\{p_1, p_2, ..., p_n\}$ imposed by the control law of the system* (16) *and* (17)*, be real, distinct and negative. The control law u is derived by fusion of the feedback signals and expressed by:*

$$u = -Lx \ , \ L = \sum_{i=1}^{r} v_i L_i \tag{18}$$

where the validity of the model M_i is v_i, $v_i \in [0,1]$, $\forall i = 1, 2, ..., N$, $\sum_{i=1}^{r} v_i = 1$, is stabilizing if the following condition is met:

$$(\left|\gamma'_{i,n}(.)\right| - \sum_{k=1}^{n-1} \frac{|\beta_k| \left|\gamma'_{i,k}(.)\right|}{p_k}) \prec 0 \tag{19}$$

with:

$$\beta_k = \prod_{\substack{j=1 \\ j\neq k}}^{n-1} (p_k - p_j)^{-1}, \quad \forall\, k = 1, \cdots, n-1 \tag{20}$$

$$\gamma'_{i,k}(.) = -P'_{A_i(.)-BL}(p_k)\ , \quad \forall\, k = 1, \cdots, n-1 \tag{21}$$

$$\gamma'_{i,n}(.) = trace\,[A_i(.) - BL(.)] - \sum_{j=1}^{n-1} p_j \tag{22}$$

Proof: Knowing that the free state matrix of the closed loop system is in a controllable form, a change of base under the form T (6), with: $\alpha_j \neq \alpha_k\,, \forall\, k \neq j$ allows the free state matrix to be in Benrejeb arrow form:

$$\begin{bmatrix} \alpha_1 & & & \beta_1 \\ & \ddots & & \vdots \\ & & \alpha_{n-1} & \beta_{n-1} \\ \gamma'_{i,1}(.) & \cdots & \gamma'_{i,n-1}(.) & \gamma'_{i,n}(.) \end{bmatrix} \tag{23}$$

Then, by applying the Borne-Gentina criterion, it comes, for $p_k = \alpha_k$ real and negative, $k = 1, \cdots, n-1$, that the studied closed loop system is stable if:

$$(-1)^n \det \begin{bmatrix} p_1 & & & |\beta_1| \\ & \ddots & & \vdots \\ & & p_{n-1} & |\beta_{n-1}| \\ \left|\gamma'_{i,1}(.)\right| & \cdots & \left|\gamma'_{i,n-1}(.)\right| & \gamma'_{i,n}(.) \end{bmatrix} > 0 \tag{24}$$

Since the new dynamic is characterized by the distinct poles imposed on the system, choosing $\alpha_j = p_j$ for $(j = 1, \cdots, n-1)$, permits us in the case when each model is chosen with A_i constant and its validity equal to one, and the other validities being null, to conclude that overall $\gamma'_{i,j}$ are null for $(j = 1, \cdots, n-1)$, and $\gamma'_{i,n} = p_n$. Under these conditions, expression (7) becomes:

$$T^{-1}\,(A_i - BL(.))\,T = \begin{bmatrix} p_1 & & & \beta_1 \\ & \ddots & & \vdots \\ & & p_{n-1} & \beta_{n-1} \\ 0 & \cdots & 0 & p_n \end{bmatrix} \tag{25}$$

and the system is stable since: $p_j < 0\,, \forall\, j = 1, ..., n$.

6 On continuous nonlinear TSK fuzzy models stability conditions

The study of the stability property of an open or a closed-loop linear or non linear plant is generally based on the assumption that mathematical model of the system under consideration is known.
For complex and large scale systems, such a mathematical model is often unknown or ill defined. The use of a fuzzy model, in this case, adds a highest level of stability study complexity of such systems and of the linear ones which are becoming non linear with the use of fuzzy approach [38].
Since Mamdani applied fuzzy logic to a practical control problem, many different fuzzy systems have been used with different structures, membership functions, etc, classified into the following three types according to their consequent parts: type of linguistic consequent parts, called Mamdani-type, type of singleton consequent parts, called singleton-type, and type of linear consequent parts named (TS). (TS) fuzzy models, proposed by Takagi and Sugeno and further developed by Sugeno and Kang (TSK) [37, 38], which are nonlinear systems described by a set of IF-THEN rules, giving a local linear representation of an underlying system, can describe or approximate a wide class of nonlinear systems. In this approach, local dynamics in different state space regions are represented by linear models and the overall system as the fuzzy interpolation of these linear models.
For closed loop system, consisting on a linear or a non linear process and a fuzzy controller, the methods known for testing the stability can be subdivided into two groups as: time domain methods: Lyapunov theory, hyperstability

theory, bifurcation theory and graph theory; and frequency domain methods as: harmonic balance, circle criterion, and Popov criterion.
The basic idea of the use of the Lyapunov theory is to define a Lyapunov function and to show with the help of the negative semi-definite derivative that equilibrium points exist. Different versions are known. One way to use this theory is to derive conditions for the fuzzy controller out of the negative semi-definite derivative of a predefined Lyapunov function and the given model of the process. Another way is to interpret the fuzzy controller as a non linear static function and to try to find an adapted Lyapunov function with the help of the method of Aizerman. A third way is to use facet functions. In this case, the fuzzy controller is realized by boxwise multilinear or polytop-wise affine facet functions and the plant is described by a state-space model. For testing the stability, a numerical parameter optimisation procedure is required. Hence, it is important to study their stability or to synthetize their stabilizing controllers.
In fact, the stability study constitutes an important phase in the synthesis of a control law, as well as in the analysis of the dynamic behavior of a closed loop system. It has been one of the central issues concerning fuzzy control, refer to the brief survey on the stability issues given in [38].
In recent literature, Tanaka and Sugeno [39], have provided a sufficient condition for the asymptotic stability of a fuzzy system in the sense of Lyapunov through the existence of a common Lyapunov function for all the subsys-tems.
This kind of design methods suffer mainly from few limitations : (1) One can construct a (TS) model if local description of the dynamical system to be controlled is available in terms of local linear models; (2) A common positive definite matrix must be found to satisfy a matrix Lyapunov equation, which can be difficult especially when the number of fuzzy rules required to give a good plant model is large so that the dimension of the matrix equation is high; (3) It appears that a necessary condition, for the existence of this common positive definite matrix, is that all subsystems must be asymptotically stable.
To overcome those difficulties, we propose, in this paper, to study the stability of (TS) fuzzy nonlinear model through the study of the convergence of a regular vector norm.
If the vector norm is of dimension one, then this is like the second Lyapunov method approach; therefore, if it is of higher dimension, then we deal with a vector-Lyapunov function [9].
The vector norm approach, based on the comparison / overvaluing principle, has a major advantage : it deals with a very large class of systems, since no restrictive assumption is made on the matrices of state equations, except that they are bounded for bounded states, in such a way that a unique continuous solution exists.
Nevertheless, although the overvaluing principle allows the simplification of the study, it also presents the corre-sponding drawback: overvaluation means losing information on the real behavior of the process. In many cases, this type of drawback can be bypassed by using changes of state variables leading to a good performance of the representation [3, 8]. For instance, for continuous control, a particularly interesting case is the one in which the off-diagonal elements are naturally positive or equal to zero; in this case, the overvaluing is carried out without loss of information.

TSK fuzzy nonlinear continuous model description

Let us cconsider a (TS) fuzzy model when local description of the process to be controlled is available in terms of nonlinear autonomous models:

$$\dot{x}(t) \;=\; A_i(x)\, x(t) \tag{26}$$

where: $x \in R^n$ describes the state vector, $A_i(.)$ are matrices of appropriate dimensions, $A_i(.) = \left\{a_{ij}(.)\right\}$ and $a_{ij}(.) \;:\; R^n \rightarrow R$, are nonlinear elements.
It is assumed that $x = 0$ is the unique equilibrium state of the studied system.
The above information is then fused with the available IF-THEN rules where the i^{th} rule, $i = 1,...,r$, can have the form: IF $\{x(t)\, is H_i(x)\}$ THEN $\{\dot{x}(t) \;=\; A_i(.)\, x(t)\}$, where $H_i(x)$ is the grade of the membership of the state $x(t)$.
The final output of the fuzzy system is inferred as follows:

$$\dot{x}(t) \;=\; \sum_{i=1}^{r} h_i(x) A_i(.)\, x(t) \tag{27}$$

with, for $i = 1,...,r$, $0 \leq h_i(x) \;\leq 1$ and $\sum\limits_{i=1}^{r} h_i(t) = 1$.
It is straightforward to show that a sufficient condition for asymptotic stability in the large of the equilibrium state $x = 0$ of the unforced fuzzy model, obtained by linearization of (26), A_i constant $\forall\, i$,

$$\dot{x}(t) \;=\; \sum_{i=1}^{r} h_i A_i x(t) \tag{28}$$

is that there exists a common symmetric positive definite matrix P such that, for $i = 1,2,...,r$:

$$A_i^T P + PA_i < 0 \tag{29}$$

The necessary condition for the existence of matrix P is that each matrix A_i must be asymptotically stable [40], i.e. all the subsystems are stable. Then, these conditions are very conservative.

TSK fuzzy nonlinear model stability condition

Let us consider the non linear continuous process (26) whose model is in the controllable form, that matrices $A_i(.)$, of equation (27), are written as:

$$A_i(.) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & & & 0 \\ 0 & & \cdots & 0 & 1 \\ -a_{i,0}(.) & & \cdots & & -a_{i,n-1}(.) \end{bmatrix} \tag{30}$$

The final description of the fuzzy system is then inferred as follows:

$$\dot{y}(t) = N(.)\,y(t) \tag{31}$$

where $y(t)$ is the new state vector such that $x(t) = Ty(t)$ and:

$$N(.) = \sum_{i=1}^{r} h_i(.)F_i(.) \tag{32}$$

$$N(.) = \begin{bmatrix} \alpha_1 & & & \beta_1 \\ & \ddots & & \vdots \\ & & \alpha_{n-1} & \beta_{n-1} \\ \sum_{i=1}^{r} h_i\gamma_{i,1}(.) & \cdots & \sum_{i=1}^{r} h_i\gamma_{i,n-1}(.) & \sum_{i=1}^{r} h_i\gamma_{i,n}(.) \end{bmatrix} \tag{33}$$

In such conditions, if $p(y)$ denotes a vector norm of y, satisfying component to component the equality: $p(y) = |y|$, it is possible, by the use of the aggregation techniques [9], to define a comparison system (13), $z \in R^n$, of (31):

$$\dot{z}(t) = M(.)\,z(t) \tag{34}$$

In this expression, the comparison matrix $M(.)$ is deduced from the matrix $N(.)$ by substituting only the off-diagonal elements by their absolute values; it can be written as:

$$M(.) = \begin{bmatrix} \alpha_1 & & & |\beta_1| \\ & \ddots & & \vdots \\ & & \alpha_{n-1} & |\beta_{n-1}| \\ \left|\sum_{i=1}^{r} h_i\gamma_{i,1}(.)\right| & \cdots & \left|\sum_{i=1}^{r} h_i\gamma_{i,n-1}(.)\right| & \sum_{i=1}^{r} h_i\gamma_{i,n}(.) \end{bmatrix} \tag{35}$$

Noting that the non-constant elements are isolated in the last row of matrix $M(.)$, then the stability condition of the continuous nonlinear system (27) can be easily deduced from the Borne-Gentina criterion [17]. It comes:

$$(-1)^i \Delta_i > 0, i = 1,2,...,n \tag{36}$$

with Δ_i the i th $M(.)$ principal minor.

It is clear that, for $i = 1,2,...,n-1$, the condition (36) is verified for $\alpha_i \in R_-$, therefore, for $i = n$ and using the relation (36), it leads to the stability condition (37).

Then, the (TS) fuzzy nonlinear model stability, in the continuous case, can be studied by the following proposed theorem.

Theorem 8. *If there exist* $\alpha_i \in R_-$, $i = 1,2,...,n$, $\alpha_i \neq \alpha_j, \forall i \neq j$ *and* $\varepsilon \in R_+$, *such that the inequality:*

$$-\sum_{i=1}^{r} h_i \gamma_{i,n}(.) + \sum_{j=1}^{n-1} \left| \sum_{i=1}^{r} h_i \gamma_{i,j}(.) \beta_j \right| \alpha_j^{-1} \geq \varepsilon \ \forall x \in R^n \tag{37}$$

is satisfied, the equilibrium state of the studied continuous non linear system (27) *and* (30) *is asymptotically globally stable.*

If there exist α_j, $j = 1,2,...,n-1$, such that:

$$\sum_{i=1}^{r} h_i \gamma_{i,j}(.) \beta_j > 0 \ j = 1,2,...,n-1 \tag{38}$$

the theorem 2 can be simplified and the comparison system (35) can be chosen identically to (33).
Since for $N(.)$:

$$\Delta_n = \sum_{i=1}^{r} h_i P_{A_i}(.0) \tag{39}$$

$$-\sum_{i=1}^{r} h_i \gamma_{i,n}(.) + \sum_{j=1}^{n-1} \alpha_j^{-1} \sum_{i=1}^{r} h_i \gamma_{i,j}(.) \beta_j = \prod_{j=1}^{n-1} (-\alpha_j)^{-1} \sum_{i=1}^{r} h_i P_{A_i}(.,0) \tag{40}$$

Hence to corollary 1.

Corollary 9. *If there exist* $\alpha_j \in R_-, \alpha_j \neq \alpha_k, \forall j \neq k$ *and* $\varepsilon \in R_+$ *such that:*

(i) the inequalities (38) are satisfied, $\forall x \in R^n$

(ii) $\sum_{i=1}^{r} h_i(t) P_{A_i}(.,0) \geq \varepsilon, \ \forall x \in R^n$ (41)

the equilibrium state of the continuous system described by (27) *and* (30) *is globally asymptotically stable.*

The obtained stability conditions, explicitly expressed by the studied models and fuzzification parameters, applicable for (TS) fuzzy models in particular, make the approach useful for the synthesis of stabilization fuzzy control law.

7 Conclusion

In this paper, the influence of the state vector description on the determination of the stability condition, based on Borne and Gentina stability criterion using vector norms and overvaluing systems is discussed for continuous non linear large scale systems.
With a description by the use of Benrejeb arrow form matrices, and of vector norms as Lyapunov functions, the criterion defines large classes of systems for which the Aizerman conjecture to a comparison system is satisfied.
The use of this approach is generalized to the stability study of multimodel system control and TSK fuzzy systems.
Other similar results can be obtained easily for nonlinear discrete large scale systems.

References

[1] J. Aracil, A. Garcia-Cezero, A. Barreiro, A. Ollero, “Fuzzy control of dynamical systems and stability analysis based on the conicity criterion,” *IFSA’91World Congress*, Brussels, July, 5 - 8, 1991.

[2] A. Barreiro, J. Aracil, “Stability of uncertain dynamical systems," *IFAC on AI in Real-Time Control*, Delft, pp. 177-182, 1992.

[3] M. Benrejeb, “Sur l’analyse et la synthèse de processus complexes hiérarchisés. Application aux systčmes singuličrement perturbés," *Doctor Science Thesis*, USTL, Lille, 1980.

[4] M. Benrejeb, M.N. Abdelkrim, “On order reduction and stabilization of TSK non linear fuzzy models by using arrow form matrix," *SAMS*, Vol. 43, n°7, pp. 977-991, 2003.

[5] M. Benrejeb, P. Borne, F. Laurent, “Sur une application de la représentation en flèche à l’analyse des processus," *RAIRO Automatique*, Vol. 16, n°2, pp. 133-146, 1982.

[6] M. Benrejeb, M. Gasmi, “On the use of an arrow form matrix for modeling and stability analysis of singularly perturbed nonlinear systems," *SAMS,* 40, pp. 209-225, 2001.

[7] M. Benrejeb, A. Sakly, K. Ben Othman, P. Borne, “Choice of conjunctive operator of TSK fuzzy systems and stability domain study," *MATCOM*, to appear, Jan. 2008.

[8] M. Benrejeb, D. Soudani, A. Sakly, P. Borne, “New discrete TSK fuzzy systems characterization and stability domain," *IJCCC*, Vol. I, n°4, pp. 9-19, 2006.

[9] P. Borne, “Non linear system stability. Vector norm approach," *System and Control Encyclopedia*, t.5, pp. 3402-3406, Pergamon Press, 1987.

[10] P. Borne, M. Benrejeb, “On the stability of a class of interconnected systems," *IFAC Symposium MVTS*, Fredericton, pp. 261-267, 1977.

[11] P. Borne. P. Vanheeghe, E. Duflos, *Automatisation des processus dans l’espace d’état*, Ed. Technip, Paris, 2007.

[12] Y.Y. Chen, “Stability analysis of fuzzy control - A Lyapunov approach," *IEEE Annual Conference on SMC*, Vol. 3, pp. 1027-1031, 1987.

[13] F. Delmote, “Analyse multimodèle," *Doctor Thesis*, USTL, Lille, 1997.

[14] D. Driankov, H. Hellendoorn, M. Reinfrank, *An Introduction to Fuzzy Control*, Springer Verlag, Berlin, Heidelberg, 1993.

[15] A. El Kamel, P. Borne, M. Ksouri-Lahmari, M. Benrejeb, “On the stability of non-linear multimodel systems," *SACTA*, Vol. 2, 1999.

[16] F.R. Gantmacher, *Théorie des matrices,* Ed. Dunod, Paris, 1966.

[17] J.C. Gentina, P. Borne, F. Laurent, “Stabilité des systèmes continus non linéaires de grande dimension," *RAIRO*, Août, J-3, pp. 69-77, 1972.

[18] J.C. Gentina, P. Borne, C. Burgat, J. Bernoussou, L.T. Grujic, “Sur la stabilité des systèmes de grande dimension. Normes vectorielles," *RAIRO Aut./Sys. Anal. and Cont.*,Vol. 13, n°1, pp. 57-75, 1979.

[19] L.T. Grujic, “Non Lyapunov stability analysis of large-scale systems on time varying sets," *Int. J. Control*, Vol. 21, n°3, pp. 401-415, 1975.

[20] L.T. Grujic, “On absolute stability and the Aizerman conjecture," *Automatica*, Vol. 17, pp. 335-349, 1981.

[21] L.T. Grujic, “General stability of large-scale systems," *IFAC on Large Scale Systems Theory and Applications*, pp. 16-20, June, Udine, 1976.

[22] L.T. Grujic, J.C. Gentina, P. Borne, “General aggregation of large scale systems by vector Lyapunov functions and vector norms," *Int. J. of Control,* Vol. 24, n°4, pp. 529-550, 1976.

[23] L.T. Grujic, J.C. Gentina, P. Borne, C. Burgat, J. Bernoussou, “Sur la stabilité des systèmes de grande dimension. Fonctions de Lyapunov vectorielles," *RAIRO Aut./Sys. Anal. and Cont.*, Vol. 12, n°4, pp. 319-348, 1978.

[24] L.T. Grujic, D.D. Siljak, “Asymptotic stability and instability of large scale systems," *IEEE Trans. on Auto. Control,* Vol. 18, n°6, Dec. 1973.

[25] W. Hahn, *Stability of the motion,* Springer Verlag, Berlin, 1967.

[26] M. Kannou, “Analyse et synthèse de processus dynamiques par approche multimodčle," *Doctor Thesis*, ENIT, Tunis, 2005.

[27] W.J. Kickert, E.H. Mamdani, “Analysis of a fuzzy logic controller," *Fuzzy Sets and Systems*, Vol. 1, pp. 29-44, 1978.

[28] H. Kiendl, “Harmonic balance for fuzzy control systems," *EUFIT*, Aachen, pp. 137-141, Sept., 1993.

[29] M. Ksouri-Lahmari, “Contribution à la commande multimodèle de processus complexes," *Doctor Thesis*, USTL, Lille, 1999.

[30] M. Ksouri-Lahmari, P. Borne, M. Benrejeb, "Multimodel: the construction of model bases," *Studies in Informatics and Control*, Vol. 3, n°3, pp. 199-210, 2004.

[31] A. M. Lyapunov, "Problème général de la stabilité du mouvement. (in French)," *Ann. Fac. Sci. Toulouse*, Vol. 9, pp. 203-474, 1907, Reprinted in *Ann. Math. Study, Princeton University Press*, n°17, 1949.

[32] V.M. Matrosov, "On the theory of stability of motion," *Prikl. Mat. Mekh.,* Vol. 26, pp. 992-1002, 1962.

[33] K.S. Ray, D.D. Majumder, "Application of circle criteria for stability analysis associated with fuzzy logic controller," *IEEE Trans. on Syst. Man and Cyber.*, Vol. 14, pp. 345-349, 1984.

[34] M. Naceur, D. Soudani, M. Benrejeb, P. Borne, "On internal multimodel control for non linear systems," *IMACS Multi-conference on Comp. Eng. In Syst. Appli.*, Beijing, pp. 306-310, 2006.

[35] F. Robert, "Normes vectorielles de vecteurs et de matrices," *R.F.T.I. Chiffres*, Vol. 17, n°4, pp. 261-299, 1964.

[36] D.D. Siljak, "Stability of large scale systems under structural perturbations," *IEEE Trans. on Syst. Man and Cyber.*, Vol. 2, n°5, Nov., 1972.

[37] M. Sugeno, G.T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets and Systems*, Vol. 28, pp. 15-33, 1988.

[38] M. Sugeno, "On stability of fuzzy systems expressed by fuzzy rules with singleton consequents," *IEEE Trans. on Fuzzy Systems*, Vol. 7, n°2, pp. 201-224, 1999.

[39] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans.on Syst. Man and Cyber.,* Vol. 15, n°1, pp. 16-132, 1985.

[40] K. Tanaka, M. Sano, "Stability analysis and design of fuzzy control systems," *Fuzzy Sets and Systems,* Vol. 45, pp. 135-156, 1992.

[41] M.C.M. Teixeira, S.H. Zak, "Stabilizing controller design for uncertain nonlinear systems using fuzzy models," *IEEE Trans. on Fuzzy Systems,* Vol. 7, n°2, pp. 133-142, 1999.

[42] H.O. Wang, K. Tanaka, M.F. Griffin, "An approach to fuzzy control of nonlinear systems: Stability and design issues," *IEEE Trans. Fuzzy Systems,* Vol. 4, pp. 14-23, 1996.

[43] L.X. Wang, "Stable adaptive fuzzy control of nonlinear systems," *IEEE Trans. on Fuzzy Systems,* Vol. 1, pp. 146-155, 1993.

[44] W. Wen-June, L. LEH, "Stability and stabilization of fuzzy large-scale systems," *IEEE Trans. on Fuzzy Syst.,* Vol. 12, n°3, pp. 309-315, 2004.

Appendix

Vector norm concept and overvaluing technique

Definition of vector norm

Let $E = R^n$ be a vector space and $E_1, E_2, ..., E_k$ subspaces of E which verify: $E = E_1 \cup E_2 \cup ... \cup E_k$. Let $x \in E$ be an n vector defined on E with a projection in the subspace E_i denoted by x_i, $x_i = P_i x$, where P_i is a projection operator from E into E_i, p_i is a scalar norm $(i = 1, 2, ..., k)$ defined on the subspace E_i and p denotes a vector norm of dimension k and with i th component, $p_i(x) = p_i(x_i)$, $p_i(x) : R^n \rightarrow R_+^k$, where $p_i(x_i)$ is a scalar norm of x_i.

Definition of an overvaluing system

The matrix $M : T \times R^n$ defines an overvaluing system (S) with respect to the vector norm p if and only if the following inequality is verified for each corresponding component: $D^+ p(x) \leq M(A(t,x))p(x) \ \ \forall x \in E \ \ \forall t \in T_0$, where D^+ denotes the right hand derivative.

Let us denote: $A_{ij}(t,x) = P_{ri}^* A(t,x) P_{rj}^*, \forall i, j = 1, 2, ..., k$; $M(t,x) = \left\{\mu_{ij}(t,x)\right\}$;

$$m_{ij}(t,x,y) = \frac{[grad\, p_i(y_i)]^T A_{ij}(t,x) y_i}{p_i(x_i)}, \ \forall i, j = 1, 2, ..., k.$$

Definition of pseudo-overvaluing matrix

The matrix $M(t,x) = \left\{\mu_{ij}(t,x)\right\}$, $\mu_{ij} : T \times R^n \rightarrow R$, is a pseudo-overvaluing matrix of $A(t,x)$ if and only if:

$\mu_{ii}(t,x) = \sup_{y\in E}\{m_{ii}(t,x,y)\}, \forall i = 1,2,...,k, \forall x \in E, \forall t \in T_0$

$\mu_{ij}(t,x) = \max(0, \sup_{y\in E}\{m_{ij}(t,x,y)\}), \ \forall i \neq j = 1,2,...,k, \ \forall t \in T_0$

It is sometimes convenient to determine a pseudo-overvaluing matrix which depends on t only or which is constant. In such a case, we have $M \geq M(t) \geq M(t,x)$.

Pierre Borne[1] and Mohamed Benrejeb[2]
[1] LAGIS, Ecole Centrale de Lille
BP 48, 59651 Villeneuve d'Ascq Cedex, France
[2]LARA, Ecole Nationale d'Ingénieurs de Tunis
BP 37, Tunis Le Belvédčre, Tunisie
E-mail: pierre.borne@ec-lille.fr ; mohamed.benrejeb@enit.rnu.tn

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 67-74

Non-negative Matrix Factorization, A New Tool for Feature Extraction: Theory and Applications

Workshop invited key lecture

Ioan Buciu

Abstract: Despite its relative novelty, non-negative matrix factorization (NMF) method knew a huge interest from the scientific community, due to its simplicity and intuitive decomposition. Plenty of applications benefited from it, including image processing (face, medical, etc.), audio data processing or text mining and decomposition. This paper briefly describes the underlaying mathematical NMF theory along with some extensions. Several relevant applications from different scientific areas are also presented. NMF shortcomings and conclusions are considered.
Keywords: Non-negative matrix factorization, image decomposition, applications.

1 Introduction

Data decomposition has multiple meanings and goals, arising from many applications, such as data compression, transmission or storage. An important application comes from the pattern recognition field, where the purpose is to automatically cluster the data samples into distinct classes. This task usually requires the extraction of discriminant latent features from the initial data prior to classification. Through the feature extraction issue the computational load is reduced and and possible discovery of task-relevant hidden variables can be performed. Feature extraction is possible as the data we perceive may contain significant redundant information. Another reason is that we do not need the full information. Rather, the extracted information is application-dependent. For example, in a facial expression recognition task, we do not care about the whole face. We can discard information related to the cheek or nose, which does not contribute to discriminate among expressions, but, certainly the eyes or mouth region encapsulate valuable information which is highly relevant for this task and can not be neglected.

Non-negative Matrix Factorization is a relatively recent approach to decompose data into two factors with non-negative entries. At least two reasons exist to constraint the entries being non-negative. The first reason comes from neurophysiology where the firing rate of visual perception neurons are non-negative. The second reason comes from the image processing field, where the intensity images have nonnegative values. Given a matrix $\mathbf{X}$ of size $m \times n$ whose columns contain data samples, the data decomposition task can be described by factoring $\mathbf{X}$ into two terms $\mathbf{W}$ and $\mathbf{H}$ of size $m \times p$ and $p \times n$, respectively, where $p <\min(m,n)$. The decomposition is performed so that the product $\mathbf{WH}$ should approximate as best as possible the original data $\mathbf{X}$, for $p <\min(m,n)$. The columns of $\mathbf{W}$ are usually called basis vectors and the rows of $\mathbf{H}$ are called decomposition (or encoding) coefficients. Thus, the original data are represented as linear combinations of these basis vectors. Contrary to other decomposition techniques which allow some factors to vanish (due to their equal values and opposite signs, as in Principal Component Analysis, for instance), NMF provides a more intuitive and meaningful decomposition allowing only additive operations.

This paper briefly describes the underlaying mathematical NMF theory along with some extensions. Several relevant applications from different scientific areas are also presented. NMF shortcomings and conclusions end the paper.

2 Non-negative Matrix Factorization Approach

Through the decomposition process, each element x_{ij} of the matrix $\mathbf{X}$ can be written as $x_{ij} \approx \sum_k w_{ik}h_{kj}$. The quality of approximation depends on the cost function used. Two cost functions were proposed by Lee and Seung in [1]: the Euclidean distance between $\mathbf{X}$ and $\mathbf{WH}$ and KL divergence. KL based cost function is expressed as:

$$D_{NMF}(\mathbf{X}\|\mathbf{WH}) \triangleq \sum_{i,j}\left(x_{ij}\ln\frac{x_{ij}}{\sum_k w_{ik}h_{kj}} + \sum_k w_{ik}h_{kj} - x_{ij}\right), \tag{1}$$

This expression can be minimized by applying multiplicative update rules subject to $\mathbf{W},\mathbf{H} \geq 0$. Since both matrices $\mathbf{W}$ and $\mathbf{H}$ are unknown, we need an algorithm which is able to find these matrices by minimizing the divergence

(1). By using an auxiliary function and the Expectation Maximization (EM) algorithm [2], the following update rule for computing h_{kj} is found to minimize the KL divergence at each iteration t [1]:

$$h_{kj}^{(t)} = h_{kj}^{(t-1)} \frac{\sum_i w_{ki}^{(t)} \frac{x_{ij}}{\sum_k w_{ik}^{(t)} h_{kj}^{(t-1)}}}{\sum_i w_{ik}^{(t)}}. \tag{2}$$

By reversing the roles of $\mathbf{W}$ and $\mathbf{H}$ in (2), a similar update rule for each element w_{ik} of $\mathbf{W}$ is obtained:

$$w_{ik}^{(t)} = w_{ik}^{(t-1)} \frac{\sum_j \frac{x_{ij}}{\sum_k w_{ik}^{(t-1)} h_{kj}^{(t)}} h_{jk}^{(t)}}{\sum_j h_{kj}^{(t)}}. \tag{3}$$

Both updating rules are applied alternatively in an EM manner and they guarantee a nonincreasing behavior of the KL divergence.

2.1 Non-negative Matrix Factorization Extensions

Several NMF variants have been proposed in the literature for tailoring the standard NMF approach to specific applications or rationales. Li et al [3] have developed a variant, termed Local Non-negative Matrix Factorization (LNMF) algorithm, imposing more constraints to the KL cost function to get more localized image features. The associated cost function is then given by:

$$f_{LNMF}^{KL}(\mathbf{X}||\mathbf{WH}) \triangleq f_{NMF}^{KL}(\mathbf{X}||\mathbf{WH}) + \alpha \sum_{ij} u_{ij} - \beta \sum_i v_{ii}, \tag{4}$$

where $[\mathbf{u}_{ij}] = \mathbf{U} = \mathbf{W}^T\mathbf{W}$ and $[\mathbf{v}_{ij}] = \mathbf{V} = \mathbf{H}\mathbf{H}^T$. Here, α and $\beta > 0$ are constants. By maximizing the third term in (4), the total squared projection coefficients over all training images is maximized. The second term can be further split into two parts:

1. $\sum_i u_{ii} \longrightarrow$ min. This term guarantees the generation of more localized features on the basis images $\mathbf{Z}$, than those resulting from NMF, since the basis image elements are constrained to be as small as possible.

2. $\sum_{i \neq j} u_{ij} \longrightarrow$ min. This enforces basis orthogonality, in order to minimize the redundancy between image bases.

The following factors updating rules were found for the KL-based LNMF cost function:

$$h_{kj}^t = \sqrt{h_{kj}^{t-1} \sum_i w_{ki} \frac{x_{ij}}{\sum_k w_{ik} h_{kj}^{t-1}}} \tag{5}$$

$$w_{ik}^t = \frac{w_{ik}^{t-1} \sum_j \frac{x_{ij}}{\sum_k w_{ik}^{t-1} h_{kj}} h_{jk}}{\sum_j h_{kj}}. \tag{6}$$

LNMF was further extended by Buciu and Pitas [4], who developed a NMF variant that takes into account class information. Their algorithm, termed Discriminant Non-negative Matrix Factorization (DNMF), leads to a class-dependent image representation. The KL-based DNMF cost function is given by:

$$f_{DNMF}^{KL}(\mathbf{X}||\mathbf{WH}) \triangleq f_{LNMF}^{KL}(\mathbf{X}||\mathbf{WH}) + \gamma \mathbf{S}_w - \delta \mathbf{S}_b, \tag{7}$$

where γ and δ are constants. The new terms are the within-class $\mathbf{S}_w$ and the between-class $\mathbf{S}_b$ scatter matrix, respectively, expressed as following:

1. $\mathbf{S}_w = \sum_{c=1}^{\mathcal{Q}} \sum_{l=1}^{n_c} (\mathbf{h}_{cl} - \mu_c)(\mathbf{h}_{cl} - \mu_c)^T$. Here, $\mathcal{Q}$ are the image classes and n_c is the number of training samples in class $c = 1, \ldots, \mathcal{Q}$. Each column of the $p \times n$ matrix $\mathbf{H}$ is viewed as image representation coefficients vector $\mathbf{h}_{cl}$, where $c = 1, \ldots, \mathcal{Q}$ and $l = 1, \ldots, n_c$. The total number of coefficient vectors is $n = \sum_{c=1}^{\mathcal{Q}} n_c$. Further, $\mu_c = \frac{1}{n_c} \sum_{l=1}^{n_c} \mathbf{h}_{cl}$ is the mean coefficient vector of class c, and $\mu = \frac{1}{n} \sum_{c=1}^{\mathcal{Q}} \sum_{l=1}^{n_c} \mathbf{h}_{cl}$ is the global mean coefficient vector.

2. $\mathbf{S}_b = \sum_{c=1}^{\mathcal{Q}} (\mu_c - \mu)(\mu_c - \mu)^T$ defines the scatter of the class mean around the global mean μ.

By imposing these terms in the cost function, the decomposition coefficients now encode class information and they are updated according to the following expression:

$$h_{kl(c)}^{(t)} = \frac{2\mu_c - 1 + \sqrt{(1-2\mu_c)^2 + 8\xi h_{kl(c)}^{(t-1)} \sum_i w_{ki}^{(t)} \frac{x_{ij}}{\sum_k w_{ik}^{(t)} h_{kl(c)}^{(t-1)}}}}{4\xi} \tag{8}$$

where $\xi = \gamma - \beta$ is a constant. The elements h_{kl} are then concatenated for all $\mathcal{Q}$ classes as:

$$h_{kj}^{(t)} = [h_{kl(1)}^{(t)} \,|\, h_{kl(2)}^{(t)} \,|\, \ldots \,|\, h_{kl(\mathcal{Q})}^{(t)}] \tag{9}$$

where "|" denotes concatenation. The expression for updating the basis vectors remains the same as in the LNMF approach.

Spareness is an important issue for image decomposition and representation in the Human Visual System (HVS). Many theoretical papers and experiments brought evidences that the response of the mammalian primary visual cortex (know also as V1 neurons) can be described by localized, oriented and bandpass filters (also known as receptive fields). When applied to natural images, these filters decompose the images into features that are very similar to those obtained by HVS receptive fields. Viewed within this light, Hoyer [5] proposed a new NMF version called Non-negative Sparse Coding (NNSC) where auxiliary constraints are used to impose factor sparseness. The sparseness measure is based on the relation between the L_1 norm and L_2 norm, i.e., sparseness($\mathbf{x}$) = $\frac{\sqrt{m} - \|\mathbf{x}\|_1/\|\mathbf{x}\|_2}{\sqrt{m}-1}$. Furthermore, a penalty term of the form $J(\mathbf{W}) = (\zeta \|resh(\mathbf{W})\|_2 - \|resh(\mathbf{W})\|_1)^2$ is introduced in the standard NMF problem, where $\zeta = \sqrt{km} - (\sqrt{km} - 1)\eta$, and $resh(.)$ is the operator which transforms a matrix into a column vector in a column-wise fashion. Here, the desired sparseness for the basis vectors is controlled by η, which can vary from 0 to 1. By replacing $\mathbf{W}$ with $\mathbf{H}$, the sparseness control can be applied to the encoding coefficients.

A sparse NMF variant called nonsmooth NMF (nsNMF) was proposed by Montano et al. in [6], which allows a controlled sparseness degree in both factors. Like LNMF and NNSC methods, nsNMF also leads to a local image decomposition. However, unlike the LNMF approach, nsNMF explicitly modifies the sparseness degree. Also, unlike NNSC, this variant applies the sparseness concept directly to the model, achieving global sparseness. Imposing sparseness in one of the NMF factors (as NNSC does) will almost certainly force smoothness in the other in an attempt to reproduce the data as best as possible. Additionally, forcing sparseness constraints on both the basis and the encoding vectors decreases the data variance explained by the model. The new variant nsNMF seems to be more robust to this effect. The nsNMF decomposition is given by $\mathbf{X} = \mathbf{WOH}$. The matrix $\mathbf{O}_{p\times p}$ is a square positive symmetric "smoothing" matrix defined as:

$$\mathbf{O} = (1-\upsilon)\mathbf{I} + \frac{\upsilon}{p}\mathbf{1}\mathbf{1}^T, \tag{10}$$

with $\mathbf{I}$ the identity matrix and $\mathbf{1}$ is a vector of ones. The parameter $0 \leq \upsilon \leq 0$ controls the extent of smoothness of the matrix operator $\mathbf{O}$. However, strong smoothing in $\mathbf{O}$ will force strong sparseness in both the basis and the encoding vectors, in order to maintain faithfulness of the model to the data. Accordingly, the parameter υ controls the model sparseness. The suitability of the proposed method over NMF, NNSC and LNMF is investigated with respect to the deterioration of the goodness of fit between the data and the model. The nsNMF model maintained almost perfect faithfulness to the data, expressed by a variance (of goodness) value greater than 99 % for a wider range of sparseness degree, compared with the other NMF variants whose variance decreases with sparseness degree modification.

Other several NMF extensions exist. Due to space limitation we only mention some of them, including: projective NMF [7], temporal NMF with spatial decorrelation constraints [8], shifted NMF [9], incremental NMF [10], sparse higher order NMF [11], and polynomial NMF [12].

3 Non-negative Matrix Factorization Applications

The standard NMF approach and its variants have been extensively used as feature extraction techniques for various applications, especially for high dimensional data analysis. The newly formed low dimensionality subspace represented by the basis vectors should capture the essential structure of the input data as best as possible. Although, theoretically, NMF could be applied to data compression, not much work was carried out in this regard.

Rather, the computer vision community focused its attention to the application of NMF to pattern recognition applications, where the extracted NMF features are subsequently clustered or classified using classifiers. The NMF applications can be characterized according to several criteria. We provide the following application classes:

- *1D signal* applications (including sounds and EEG data), where the input matrix $\mathbf{X}$ contains in its columns one-dimensional data varying over time.
- *2D signal* applications (face object images, etc.), where the input matrix $\mathbf{X}$ contains in its columns a vectorized version of the 2D signals (basically 2D images) obtained by lexicographically concatenating the rows of the two-dimensional images.
- *Other applications*, including text or e-mail classification.

3.1 1D signal applications

The separation of pitched musical instruments and drums from polyphonic music is one application, where NMF was considered by Helén and Virtanen in [13]. The NMF splits the input data spectrogram into components which are further classified by an SVM to be associated to either pitched instruments or drums. Within this application, each column of the input matrix $\mathbf{X}$ represents a short-time spectrum vector $\mathbf{x}_t$. The non-negative decomposition takes the form $\mathbf{x}_t = \sum_{i=1}^{n} \mathbf{s}_n a_{i,t}$, where $\mathbf{s}_n$ is the spectrum of $n-th$ component, $a_{i,t}$ is the gain of $n-th$ component in frame t, and n is the component number. Individual musical instrument sounds extraction using NMF was exploited by Benetos et al. in [14]. A number of 300 audio files are used, corresponding to 6 different instrument classes (piano, violin, cello, flute, bassoon, and soprano saxophone) [15]. Two sorts of features are used to form the input matrix. The first feature set is composed of 9 audio specific features and MPEG-7 specifications (such as zero-crossing rate, delta spectrum, mel-frequency cepstral coefficients, etc). The second feature set is given by the rhythm pattern described by several other audio characteristics (such as power spectrum, critical bands, modulation amplitude, etc).

One particular NMF application is on spectral data analysis. Source spectra separation from magnetic resonance (MR) chemical shift imaging (CSI) of human brain using constrained NMF analysis was investigated by Sajda et al. [16]. In CSI, each tissue is characterized by a spectral profile or a set of profiles corresponding to the chemical composition of the tissue. In tumors, for instance, metabolites are heterogeneously distributed and, in a given voxel, multiple metabolites and tissue types may be present, so that the observed spectra are a combination of different constituent spectra. Consequently, the spectral amplitudes of the different coherent resonators are additive, making the application of NMF reasonable. The overall gain with which a tissue type contributes to this addition is proportional to its concentration in each voxel, such that $\mathbf{X}$ is the observed spectra, the columns of $\mathbf{W}$ represents concentration of the constituent materials, and the rows of $\mathbf{H}$ comprises their corresponding spectra.

The spectral data analysis was also investigated in [17] by proposing a constraint NMF algorithm to extract spectral reflectance data from a mixture for space object identification. In this application, each observation of an object is stored as a column of a spectral trace matrix $\mathbf{X}$, while its rows correspond to different wavelengths. Each column of $\mathbf{W}$, called endmember, is a vector containing nonnegative spectral measurements along p different spectral bands, where each row of $\mathbf{H}$ comprises the fractional concentration.

Multichannel EEG signals have been analyzed via NMF concept by Rutkowski et al. in [18]. The signals are firstly decomposed into intrinsic modes in order to represent them as a superposition of components with well defined instantaneous frequencies called IMF. The resulting trace IMF components form the input matrix $\mathbf{X}$, while $\mathbf{W}$ is the mixing matrix containing the true sub-spectra.

3.2 Image applications

One of the first NMF applications in images is on face recognition tasks. Li et al [3] explored this issue for both NMF and LNMF techniques, when a simple Euclidean distance is used as classifier. Their experiments revealed the superiority of LNMF over the standard NMF for the ORL face database [19], especially for occluded faces. Guillamet and Vitrià [20] also applied NMF to a face recognition task. A third framework dealing with the face recognition task is described in [21], where the DNMF is employed along NMF and LNMF for comparison. Also, besides the Euclidean distance, two other classifiers (cosine similarity measure and SVMs) are utilized. Two databases, namely ORL and YALE [22] are utilized here. The experiments showed that the NMF seems to be more robust to illumination changes than LNMF and DNMF, since the variation of illumination conditions for the faces pertaining to Yale database is much more intense than for images from the ORL database. Contrary to the

results obtained for ORL, where LNMF gave the highest recognition rate, when face recognition is performed on the YALE database, the best results are obtained by the NMF algorithm. Although the ORL database, generally, contains frontal faces or slightly rotated facial poses. This can contribute to the superior performance of LNMF, since this algorithm is rotation invariant (up to some degree), because it generates local features in contrast to NMF which yields more distributed features.

Buciu and Pitas applied NMF and LNMF for facial expression recognition in [23] and compared them with the DNMF algorithm in [4] for the same task. It was found that, for the facial expression recognition task, the DNMF method outperforms the other two techniques for the Cohn-Kanade AU-coded facial expression database [24].

The NMF for object recognition is investigated by Spratling [25], where an empirical investigation of the NMF performance with respect to its sparseness issue for occluded images is reported. The experiments were conducted for the standard bars problem, where the training data consists of 8×8 pixel images in which each of the 16 possible (one-pixel wide) horizontal and vertical bars can be present with a probability of $1/8$. The occlusion was simulated by overlapping between horizontal and vertical bars. Several NMF variants (i.e., NMF, LNMF, and NNSC) have been tested. It was found that no NMF method was able to identify all the components in the unequal bars (occlusion) problem for any value of the sparseness parameter. To overcome this situation, the author proposed a non-negative dendritic inhibition neural network, where the neural activations identified in the rows of $\mathbf{H}$ reconstruct the input patterns $\mathbf{X}$ via a set of feedforward (recognition) weights $\mathbf{W}$. When applied to face images, the proposed NMF neural network learns representations of elementary image features more accurately than other NMF variants. Guillamet et al. experimentally compared NMF to the Principal Component Analysis (PCA) for image patch classification in [20]. In these experiments, 932 color images from the Corel Image database were used. Each of these images belongs to one of 10 different classes of image patches (clouds, grass, ice, leaves, rocky mountains, sand, sky, snow mountains, trees and water). NMF outperformed PCA. Finally, a face detection approach based on LNMF was proposed by Chen et al. in [26].

Genomic signal processing is another task were NMF recently done a good job. The approach has been used to discover metagenes and molecular patterns in [27]. NMF recovered meaningful biological information from cancer-related microarray data. NMF appears to have advantages over other methods such as hierarchical clustering or self-organizing maps. NMF was found less sensitive to a priori selection of genes or initial conditions and able to detect alternative or context-dependent patterns of gene expression in complex biological systems. This ability, similar to semantic polysemy in text, provides a general method for robust molecular pattern discovery. More recent, a robust NMF variant was proposed in [28]. A least-square NMF which incorporates uncertainty estimates is developed in [29], while factoring gene expressions using NMF was also utilized in [30] where a statistical sparse NMF was developed.

3.3 Other applications

The application of NMF for text classification was undertaken in [31]. This application is characterized by a large number of classes and a small training data size. In their formulation, the elements $w_{ik} \geq 0$ represent the confidence score of assigning the k-th class label to the i-th example. Furthermore, $\mathbf{H} = \mathbf{B}\mathbf{W}^T$, where the non-negative matrix $\mathbf{B}$ captures the correlation (similarity) among different classes.

The extraction and detection of concepts or topics from electronic mail messages is a NMF application proposed by Berry and Browne in [32]. The input matrix $\mathbf{X}$ contains n messages indexed by m terms (or keywords). Each matrix element $x_{i,j}$ defines a weighted frequency at which the term i occurs in message j. Furthermore, $x_{i,j}$ is decomposed as $x_{i,j} = l_{i,j} g_i d_j$, where $l_{i,j}$ is the local weight for the term i occurring in message j, g_i is the global weight for i - th term in the subset, and d_j is a document normalization factor, which specifies whether or not the columns of $\mathbf{X}$ (i.e., the documents) are normalized. Next, a normalized term $p_{i,j} = f_{i,j} / \sum_j f_{i,j}$ is defined, where $f_{i,j}$ denotes frequency that term i appears in the message j. Then, two possible definitions exist for $x_{i,j}$. The first one sets $l_{i,j} = f_{i,j}$, $g_{i,j} = 1$, while the second interpretation sets $l_{i,j} = \log(1 + f_{i,j})$ and $g_{i,j} = 1 + (\sum_j p_{i,j} \log(p_{i,j}) / \log n)$, respectively. After NMF decomposition, the semantic feature represented by a given basis vector $\mathbf{w}_k$ (k - th column of the matrix) by simply sorting (in descending order) its i elements and generating a list of the corresponding dominant terms (or keywords) for that feature. In turn, a given row of $\mathbf{H}$ having n elements can be used to reveal messages sharing common basis vectors $\mathbf{w}_k$, i.e., similar semantic features or meaning. The columns of $\mathbf{H}$ are the projections of the columns (messages) of $\mathbf{X}$ onto the basis spanned by the columns of $\mathbf{W}$.

A chemometric application of the NMF method is proposed by Li et al. [33] where several NMF variants are used to detect chemical compounds from a chemical substances represented through Raman spectroscopy. The input matrix contains the observed total mixture chemical spectra, the basis vectors denote the contribution of chemical spectra, while the spectra is encoded into $\mathbf{H}$.

4 NMF Shortcomings

As far as the NMF open problems are concerned, several challenges exist, as follows:

- The optimization problem. All the described NMF variants suffer from the same drawback: no global minimum is guaranteed; they only lead to a local minimum, thus several algorithm runs may be necessary to avoid getting stuck in an undesired local minimum. Having an approach which conducts to a global minimum will greatly improve the numerical NMF stability.
- Initialization of **H** and **W**. Basically, the factors are initialized with random nonnegative values. A few efforts were undertaken in order to speed up the convergence of the standard NMF. Wild [34] proposed a spherical k-means clustering to initialize **W**. More recently, Boutsidis and E. Gallopoulos [35] employed an SVD-based initialization, while Buciu et al. [36] constructed initial basis vectors, whose values are not randomly chosen but contain information taken from the original database. However, this issue is an open problem and needs further improvements for the standard NMF approach and it variants.
- Subspace selection. To date, there is no approach suggesting, a priori, the optimal choice of p for the best performances. The issue is difficult and data-dependent. Typically, the algorithms run for several values of p and the subspace dimension corresponding to the highest recognition rate is reported. Also, before data projection, the resulting basis vectors may be re-ordered according to some criteria (descending order of sparseness degree, discriminative capabilities, etc).
- Nonlinear nonnegative features. Standard NMF linearly decomposes the data. The kernel-based NMF approach proposed in [12] tends to retrieve nonlinear negative features.

References

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[2] A. P. Dempster and N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data using the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.

[3] S. Z. Li, X. W. Hou and H. J. Zhang, "Learning spatially localized, parts-based representation," *Int. Conf. Computer Vision and Pattern Recognition*, pp. 207–212, 2001.

[4] I. Buciu and I. Pitas, "A new sparse image representation algorithm applied to facial expression recognition," in Proc. *IEEE Workshop on Machine Learning for Signal Processing*, pp. 539–548, 2004.

[5] P. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[6] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, nr. 3, pp. 403–415, 2006.

[7] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," *14th Scandinavian Conference on Image Analysis*, pp. 333–342, 2005.

[8] Z. Chen and A. Cichocki, "Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints," Preprint, 2005.

[9] M. Mørup, K. H. Madsen, and L. K. Hansen, "Shifted Non-negative Matrix Factorization," *IEEE Workshop on Machine Learning for Signal Processing*, 2007.

[10] S. S. Bucak, B. Gunsel, and O. Gursoy, "Incremental non-negative matrix factorization for dynamic background modelling," *ICEIS 8th International Workshop on Pattern Recognition in Information Systems*, 2007.

[11] M. Mørup, L. K. Hansen, and S. M. Arnfred, "Algorithms for sparse higher order non-negative matrix factorization (HONMF)," *Technical Report*, 2006.

[12] I. Buciu, N. Nikolaidis, and I. Pitas, "Non-negative matrix factorization in polynomial feature space," *IEEE Trans. on Neural Nerworks*, in Press, 2008.

[13] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," *13th European Signal Processing Conference*, 2005.

[14] E. Benetos, C. Kotropoulos, T. Lidy, A. Rauber, "Testing supervised classifiers based on non-negative matrix factorization to musical instrument classification," in *Proc. of the 14th European Signal Processing Conference*, 2006.

[15] Univ. of Iowa Musical Instrument Sample Database, http://theremin.music.uiowa.edu/index.html.

[16] P. Sajda, S. Du, T. Brown, R. Stoyanova, D. Shungu, X. Mao, and L. Parra, "Non-negative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. on Medical Imaging*, vol. 23, no. 12, pp. 1453–1465, 2004.

[17] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative Matrix Factorization for Spectral Data Analysis," *Linear Algebra and its Applications*, vol. 416, pp. 29–47, 2006.

[18] T. M. Rutkowski, R. Zdunek, and A. Cichocki, "Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support," *International Congress Series*, vol. 1301, pp. 266–269, 2007.

[19] http://www.uk.research.att.com/

[20] D. Guillamet and Jordi Vitrià, "Non-negative matrix factorization for face recognition," *Topics in Artificial Intelligence*, Springer Verlag Series: Lecture Notes in Artificial Intelligence, vol. 2504, pp. 336–344, 2002.

[21] I. Buciu, N. Nikolaidis, and I. Pitas, "A comparative study of NMF, DNMF, and LNMF algorithms applied for face recognition," *2006 Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, 2006.

[22] http://cvc.yale.edu

[23] I. Buciu and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," *International Conference on Pattern Recognition*, pp. 288–291, 2004.

[24] T. Kanade, J. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," *in Proc. IEEE Inter. Conf. on Face and Gesture Recognition*, pp. 46–53, 2000.

[25] M. W. Spratling, "Learning image components for object recognition," *Journal of Machine Learning Research*, vol. 7, pp. 793–815, 2006.

[26] X. Chen, L. Gu, S. Z. Li, and H.-J. Zhang, "Learning representative local features for face detection," *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1126–1131, 2001.

[27] J. P. Brunet, P. Tamayo, T.R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization", *Proc Natl Acad Sci U S A.*, vol. 101, no. 12, pp. 4164–4169, 2004.

[28] P. Fogel, S. S. Young, D. M. Hawkins, and N. Ledirac, "Inferential, robust non-negative matrix factorization analysis of microarray data", *Bioinformatics*, vol. 23, no. 1, pp.44 – 49, 2007.

[29] G. Wang, A. V. Kossenkov, and M. F. Ochs, "LS-NMF: a modified non-negative matrix factorization algorithm utilizing uncertainty estimates", *BMC Bioinformatics*, vol. 7, no. 1, pp. 175, 2006.

[30] D. Dueck, Q. D. Morris and B. J. Frey, "Multi-way clustering of microarray data using probabilistic sparse matrix factorization", *Bioinformatics*, vol. 21, no. 1, pp. 144–151, 2005.

[31] Y. Liu, R. Jin, and L. Yang, "Semi-supervised Multi-label Learning by Constrained Non-negative Matrix Factorization," in *Proc. of The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, vol. 21, pp. 421–426, 2006.

[32] M. W. Berry and M. Browne, "Email Surveillance Using Nonnegative Matrix Factorization," *Computational & Mathematical Organization Theory*, vol. 11, pp. 249–264, 2005.

[33] H. Li, T. Adali, W. Wang, D. Emge, A. Cichocki, "Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy," *Journal of VLSI Signal Processing*, vol. 48, no. 1-2, pp. 83–97, 2007.

[34] S. Wild, "Seeding non-negative matrix factorization with the spherical k-means clustering," Master thesis, University of Colorado, 2002.

[35] C. Boutsidis and E. Gallopoulos, "On SVD-based initialization for nonnegative matrix factorization," Tech. Rep. HPCLAB-SCG-6/08-05, University of Patras, Patras, Greece., 2005.

[36] I. Buciu, N. Nikolaidis, and I. Pitas, "On the initialization of the DNMF algorithm," *IEEE International Symposium on Circuits and Systems*, pp. 4671–4674, 2006.

Ioan Buciu
University of Oradea
Faculty of Electrical Engineering and Information Technology
Department of Electronics
Address: Universitatii 1, 410087, Oradea, Romania
http://webhost.uoradea.ro/ibuciu/
E-mail: ibuciu@uoradea.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 75-81

Knowledge Management using Intranets and Enterprise Portals

Parallel session invited paper

Marius Guran

Abstract: Modern companies migrate to responsive e-business models, by investing in ***information and knowledge-driven applications***, that help them respond rapidly to changing market conditions and customer needs, under the impact of globalizations and the new information and communication technologies. The knowledge and data/information harvesting is mainly customer-centric, personalization, and customization, which implies having the means to tailor the content, format, and medium of key decision-support information and knowledge to the needs of individual users. This trend is based on the use of technologies that enable the delivery of personalized information and knowledge to large number of end-users through a variety of channels, mainly internet/extranet/intranet oriented.
Many organizations have developed intranets and have understood the potential that this technology has in supporting the aims of knowledge management. In the paper is described a new vision for a ***knowledge-enabled intranet***, and is outlined how this can be achieved by using the concept of developing knowledge content artefacts. Also is presented a better way to address employees personal and job needs, by delivering more self-service capabilities and personalisation, developing ***enterprise portals***.
Keywords: intranet, knowledge-enabled intranet, knowledge management, enterprise portals, enterprise taxonomy.

1 Introduction

Traditional applications in the enterprises, mainly related to **ERP** (Enterprise Resource Planning) and **CRM** (Customer Relationship Management), are using massive amount of data on operation and customers, that are unused in datawarehouses. To turn that stored data into valuable information, companies are now questing knowledge applications (**KApps**). The business advantage in having Kapps, lies in the ability to analyze large amounts of data from any business model, determine the personalized preferences of all potentially customers, than rich them with relevant information, wherever they may be. These serves as the driving force for new generation of applications. Traditionaly, we have query-and-response paradigm for applications. For the new generation of applications, the logic is reversed: *what-if-system* didn't wait for the end user to have the question, and the system just asked the question for the end-users and send them the answer. In this way one could anticipate a whole set of questions. This new class of applications allows companies not only to collect but to analyze data and information, in order to devlope better supplier and customer relationships. It is aimed at increasing profitability through revenue grouth. This revenue-enhancing framework, focuses on an interesting mix of modeling, data processing as decision support, information retrieval, reporting and analysis, what-if-scenarios, datawarehouses, and data mining. Knowledge-driven applications have the potential to expand the use of information, by transforming existing huge data collections into revenue-generating asset [2].

2 Emerging classes of KApps

To take the full advantages for knowledge and information-based business models, there is a need for an integration framework that can tie together the various classes of Kapps. Some of the emerging classes of Kapps are [10]:

- **Customer Relationship KApps**: offer companies tools for mining customer data and information, having as outcome of this data mining process improved pricing, greater market share, longer customer retention, or a new revenue flow. For this, the companies must to do more real-time relationship management, the trend known *as personalization* (better understand and respond to each customer's needs, behavior and intentions, ensuring that customers get exactly what they need-when they need it).

- **Supply Chain KApps**: encourage trading partners to improve profits by managing inventories in the supply chain, by obtaining the information that enables visibility and certainty, offering more favorable terms, increased levels of supplies, invest in co-marketing.

- **Knowledge / Innovation Management**: assure the companies to push technologies farther, giving their employees instant access to informations and reports that previously took days or week to obtain.

- **Remote Performance Monitoring**: provide information to operating managers throughout an enterprise, that enables them to improve performance on an routine basis, by bridging operations and strategy, using key performance indicators.

- **Simulation by using *what-if scenario analysis***: encompasses advanced simulation and scenario modeling, based on information from diverse internal and external sources. This enable management to participate in developing strategies and learn risk management(by modeling of future risk and returns).

3 Architectural framework for KApps

To create an integrated decision framework, the organizations have to implement a number of KApps builted on a platform that is composed of three layers [6]:

1. **e-Business decision-support solutions**, that includes the ability to deliver views and queuing, reporting, and modeling capabilities that go beyond current offerings.

2. **Enabling technologies**: data mining, query processing, and result distribution infrastructure, that means the ability to store data in a multidimensional cube format (On-Line Analytical Processing - **OLAP**), to enable rapid data aggregation and profound analysis.

3. **Core technologies**, as data warehousing, and data markets, that get all company data working together so that user can see more, learn more, and make the organization to work better.

Because information access and control drive business competition, it is obvious to consider the lack of boundaries in modern business and that fact that corporations and consumers are becoming more interconnected via private networks and Internet. These increasing interconnections are facilitating development of KApps in three phases:

1. **Corporate Intranets**, in which the companies are creating complete and uniform linkage of information and knowledge resources distributed through the organization. For the knowledge creation to occur, data aggregation needs to be complemented with data analysis. Moving from departmental solution, in which data and reports are developed for small, specialized communities of users, to a corporate intranets, opens up data resources to a broader base of users, by using the browser as a standard interface.

2. **Extranets**, that are focusing on supply chain partners, in the conditions when the companies are moving parts of the internal corporate information infrastructure, so that suppliers and trading partners can access them(through fire-walls). The key business drivers are : fast access, customized data, and responsivennes. Standardized reports and interfaces are minimizing services requirements imposed by the management of huge data volumes, cross-platform coverage and support, response time speed, and a broad range of interface choices.

3. **Commercial Internet Applications** that focuses on new business models, created for capturing, consolidating, and reselling consumer informations, business transaction records, and financial data.

At the present, most companies and corporate strategy are in phase I, with the emphasis on creating the ability to imitate decision making through all levels of an organization. But they are facing the challenges of performing complex computational analysis on collected data and of disseminating the information and knowledge not only to employees, but also to customers, suppliers, and business partners.

4 The knowledge-enabled intranet

Most major organisations today have an intranet or several intranets, and this technology is delivering significant business benefits [1, 2, 3, 4, 7, 8, 9]. The big issue now is how to turn the intranet into the tool that it has as potential to do with knowledge-enablement than technology. The quality varies greatly from the content as limited amount of static information only, to carry a wider range of more focused information, and to be used by a growing audience.

At the first stage, the corporate intranet, as a network of the existing local intranets the questions raised refer about what information should be available, bringing that information in some defined standards for content (ex. web page for every department). The most advanced intranets usually also provide some level of paperless administration functionality (on-line services as personal records updating, expenses submissions, stationary ordering etc), saving money in internal costs. However beneficial they may be, the most advanced intranets are essentially becoming administrative tools rather than strategic business initiatives, by a full role in supporting the creation, sharing and application of the knowledge that is the core, value-creating competence of the organization. This is the meaning of the intranets to becoming knowledge-enabled. So, we can define the ***knowledge-enabled intranet (KEI)*** as an intranet that helps the organization to develop and profit from its unique knowledge assets, and supports the needs of staff in their roles as knowledge workers.

The fully **KEI**s goes further than providing read-only access to static information or paperless administrative functionality, in fulfilling corporate strategy, in supporting business processes and knowledge workers, and in developing the organization's intellectual capital, by providing a lot of facilities [6, 8, 9]:

- Focus for creation, capture, sharing, application and exploitation of the organization's differential knowledge, experience and learning.
- Support and promote efficiency, effectiveness and competitive edge through close integration with the core, knowledge-based processes and activities of organization.
- Help the people to find easily the help they need, and other people applications or documents, internal or external of the organization.
- Help people to communicate and collaborate, real-time, on-line or off-line, any time, anyplace, and anywhere.
- Automate intelligent, decision-making tasks and workflows.

From this requirements and facilities, all **KEI** need to succeed on three directions:

- be technically good and well-delivered (working in technical sense) ;
- be relevant and usable by the intended audiences and closely aligned to business processes (working in operational sense) ;
- focus on the knowledge sensitive areas (working in the business sense, supporting business strategy).

The intranet cannot become a successful knowledge management tool by adopting and upgrading sophisticated technology if is not adopted a broad approach of issues from all viewpoint: business, users, processes and content, and as well as technology. Usualy we can adopt some *key areas* to have knowledge-enabled intranet [8]:

- *Business integration*, to ensure that **KEI** address the real business priorities of the organization, supporting core competencies and strategy.
- *Process integration*, to provide optimum support at the operational level, assuring a close integration between people, tasks, workflow and content, enabling a knowledge-directed approach for use, learn, and share within each task to be mapped and analysed (ex. : a knowledge-enabled call centre).
- *Cultural alignment*, by re-organization, re-design of the jobs, by training and changes to recognition and rewards.
- *Content management*, by assuring content quality as a major issue, implementing basic processes for content sourcing, aggregation, training, sorting, editing, publishing and control.
- *New technology*, that can play a valuable role in enhancing the support the support that the intranet can provide: workflow and document management, portals, use profiling and intelligent agents.

5 Knowledge content artefacts

As presented above, a major issue and one vital way in which to knowledge-enabled an intranet, is to increase its knowledge content, qualitatively different to the kinds of content usually carried by either the static information or the paperless administration intranet. The following knowledge content artefacts might form a valuable part for the **KEI**:

- *Frequently-asked questions*, together with the answers.
- *Lessons learned* from the experience (distilled hits and tips).
- *Best practices and most efficient methods*.
- *Key contacts*, who to contact in specific situations.
- *Reminders*, as top-level warnings or suggestions.
- *Competitive analysis*, as distillation on their relative strengths and weaknesses.
- *Internal process models*, illustrating how various operating run.
- *Corporate history*, as past events and what have learned from it.
- *Instructions and procedures* needed in different occasions ("how to..")
- *Problems and solutions*, pooling the collective experience of dealing with specific difficulties.
- *Knowledge execution system*, as software applications that embody the models and instructions (programmes) to advice or guide a user.

These content artefacts differ from the content types more usually found on intranets (news, charts, documents etc), and what they have in common is that:

- they are the distilled results of some analysis, aggregation or judgement made about a group of information or a number of specific experiences or ideas;
- to function properly they need to maintain their contact with experience and analysis so that feedback from usage and new ideas can be incorporated on an ongoing basis.

This means that they are liked to the learning process of applying the knowledge in every domain, reviewing outcomes and new experience, and updating the artefacts. Such kind of artefacts can become a uniquely differential source of advice, guidance and insight that can inform and improve strategic decision-making as well as current operations.

6 Intranets and enterprise portals

Usually, employees easily accept and adopt intranets as a way of doing business. The next step is to find a better way to address employees' personal and job needs by delivering more self-service capabilities and personalisation [4, 6, 7]. For that the enterprise must have search capability as cornerstone in helping employees find available intranet-based content. The search capability had to be matched with appropriate corporate taxonomy and metadata to allow employees to browse for quality content. Enterprise taxonomies are used to describe the content that is generated by a business, and they are designed by looking at content and talking to subject matter experts so that an appropriate model of the data can be created.

Enterprise taxonomy is a multi-purpose, hierarchical list of terms that describes content, centrally managed or distributed with strong control model, that provides following benefits:

- can describe content across applications;
- are focus-agnostic (they have no explicit point of view);
- can be used to control the values in metadata;

- can extend search queries by adding synonyms, acronyms, and abbreviations, assisting site navigation.

Enterprise taxonomies are also being used to enhance corporate search, by using new search techniques, like semantic and actionable search, that are greatly improved by adding knowledge that is already built into an enterprise taxonomy structure.

A solution to enterprise/corporate intranet problems(chaos) is established to be ***enterprise portal*(EP)**, that provide a personalized window into enterprise for individual users or classes of users, based on job functions, roles, or other criteria. **EP** offers "one-stop-shopping" for knowledge workers, because is both a gateway to and a destination on the enterprise network that provides transparent, tailored access to distributed digital resources. In this way, **EP** improves efficiency and empowers employees to complete their jobs faster, better, and more effectively, increasing revenue and reducing operating costs.

Based on web-browser interface in the 4-tier web development/integration using an application and datawarehouse servers, enterprise portals can provide numerous benefits [4, 5]:

- interact with relevant information, knowledge or content and applications, both internal (via intranet) and external (via extranet, internet) to the company ;

- define business processes, including workflow-enabled processes that information automatically among various components ;

- collaborate with others (customers, suppliers, partners), both inside and outside the organization through self-service publishing ;

- provide access to business-intelligence functions, for key business events or parameters, for persons that must be alerted or notified.

Software product on the market (Oracle, Microsoft, **IBM** etc) incorporate several key feature necessary for providing a platform for enterprise-portal deployments:

1. Integration of enterprise information and application through an open interface, the portlet framework, that can be registered, and after they are available to users who have the correct access rights.

2. Incorporation of value-added core portal services, such as workflow, search capability, security, single sign-on, and an application-development environment.

3. Extensive customization and personalization capabilities at the enterprise, workgroup, and individual-user levels.

This kind of software provide a portal frame work for registration and integration of intranet-based applications and data/information stores, a set of feature-rich portal services, development tools, and a user interface that allows users to personalize their portal experiences. Some vendors are presenting knowledge management, business-intelligence, document management, or collaboration tools, as enterprise-portal products. However, these are specialized solutions that cannot meet the breadth and depth of companies' needs, especially those moving to an e-business model.

7 Intranet use for knowledge management

The enterprise portal drives integration with other applications, delivering not just the ability to find information, but to use tools to conduct the business, and allow information to be placed into applications and documents. Some examples for manufacturing companies are clarifying the role of enterprise portals and intranets for knowledge management:

- ***The Enterprise Knowledge Base*** (EKB) is an area where anyone can add to the repository, because the mission of EKB is to provide management and protection of company's informational assets, while capturing knowledge and providing large accessibility. The validity of the content and update are the responsibility of the submitter. Uers can come in and search by keyword or against taxonomies and meta-data that is a requirement for submission and approval. Frequent users may establish search profiles for re-use when visiting EKB.

- ***Knowledge Base Engineering*** (KBE) guides the design and manufacturer engineers who exclusively use CAD/CAM for their jobs. Depending on the component as system being designed/manufactured, the CAD/-CAM system contains rules, data, and dimensions that are the result of proven design or manufacturers processes. If the engineer attempt to violate a rule, for ex., the system makes suggestions based on proven knowledge and rules, allowing a rule to be violated only for innovations by flaggeding that new rules, that must have executive approval (sign-off).

- ***Best Practice Replication*** (BPR) is considered a supporting tool and process that is used primarily by members of specific communities of practice. The BPR system prompts the members (called 'focal points') via e-mail notification whenever are new best practices to review or when some time limits for responses have been expired. BPR applications can be accessed from the portal search engine specific keywords. Typically, the focal points are a conduit of knowledge, and they are responsible initially reviewing the knowledge as presented, then collaborating with their colleagues to determine if the practice is replicable at their location. The focal point then re-access the system, providing the feedback status for their location that can be :

  - **A** : adopt or adapt the practice or portions of it;
  - **NA** : not applicable practice;
  - **NEF** : not economically feasible, and the return is not worth the investment;
  - **P** : previously implemented;
  - **C** : previously adopted and now has been completed;
  - **INV** : under investigation while collaborating ($<$ 60 days).

The feedback form is designed to support project management (cost, timing, responsibilities).

8 Conclusions

The paper presented the concept of knowledge applications. and key area to address in or to have knowledge-enable intranet: business integration, process integration, cultural alignment, content management and new technologies. Also is considered the knowledge content artefact that might form a valuable part of knowledge-enable intranet , and a uniquely differential source of advice, guidance and insight that can inform and improve strategic decision-making as well as current operations.

After intranets adoption, the enterprise must have search capabilities to help employees and others to find available intranet-based content, by using enterprise portals. Enterprise portals can improve efficiency and empowers users as "focal points" to complete their jobs faster, better, and move effectively by implementing a platform for enterprise portals deployments as part of knowledge-enabled intranet.

References

[1] M. Guran, C.E. Cotet, *Manufacturing and Network Technologies*, Proceedings of FAIM 12th International Conference on Flexible Automation and Intelligent Manufacturing. William G. Sullivan, Munir Ahmad, Dieter Fichtner, Wilfried Sauer, Gerald Weigert, Thomas Zerna (Ed.), ISBN 3-486-27036-2, Dresden, Germany, pp. 36-44, July 2002.

[2] Marius Guran, *Trends and Approaches in Management of Virtual Enterprises*. Vol. Proceedings of the 15-th International Conference on Manufacturing Systems (ICMaS). Ed. Academiei Romane, Oct., 2006.

[3] M. Guran, C.E. Cotet, G. Dragoi, *The PREMINV Platform for Training in Managemen and Engineering for Virtual Enterprise*. Vol. Collaborative Business Ecosystems and Virtual Enterprises, series: IFIP (International Federation for Information Processing), Ed. Luis M. Camarinha - Matos. ISBN 1-4020-7020-9. Kluwer Academic Publisher, Boston/ London, pp. 571-579, 2002.

[4] M. Guran, C.E. Cotet, *Virtual Enterprise Environment - An Internet-Intranet-Extranet Based Tool for Customer Focusing on Quality*. Vol. Proceedings of the International Conference on Manufacturing Systems (ICMaS), Ed. Academiei Romane pp. 599-603, 2002.

[5] Marius Guran, Sisteme informationale pentru management (Cap.9: Mediul de afaceri modern si tehnologiile Internet). Curs pentru programele de Master si Doctorat la Fac. IMST, FILS (engleza), Automatica si Calculatoare, Univ. Politehnica Bucuresti, 2007.

[6] Marius Guran, Mehana Aref, Sebastian Rosu, Knowledge and Data Management in Business Intelligence. *Vol. Proceedings of 16-th International Conference on Control Systems and Computer Science*, Ed.Printech, ISBN 978-973-718-741-3, pp.39-43, May 2007.

[7] M. Guran, D. Cartu, St. Cojanu, Shared Resourses Service Center for Informatics and Management. *Vol. Proceedings of the 4-th International Symposion on Economic Informatics*. Inforec Pr. House, Bucharest, May 1999.

[8] Robert Taylor: WEB ONLY ARTICLE. InsideKnowledge, Vol. 3, Issue 8, ikmagazine, May 2000.

[9] Kwiecien Stan, Buckley Trish: From Intranet to Corporate Portals.InsideKnowledge, Vol. 6,Issue 4, ikmagazine, December 2002.

[10] Kalakota Ravi, Robinson Marcia, *e-Business Road Map for Success*, Ed. Addison-Wesley, 1999.

Marius Guran
University POLITEHNICA of Bucharest
E-mail: mguran@mix.mmi.pub.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 82-91

Cybernetics and its Romanian Forerunners

Parallel session invited paper

Ştefan Iancu

Abstract: The paper introduces the reader to the emergence of cybernetics as a science and tells the stories of the three Romanian forerunners of cybernetics: Spiru Haret, Daniel Danielopolu and Paul Postelnicu. Also here, the life and selected works of Ştefan Odobleja are presented,
Odobleja's and Wiener's cybernetic conceptions being compared. The author pointed out that Ştefan Odobleja's biggest merit is the one of having discovered the general character of the feedback and of having tried to emphasize it in the most diverse range of processes and phenomena. N. Wiener regained what, in other conditions Ştefan Odobleja has discovered, and managed to build cybernetics as a science through a complete mathematical analysis of the feedback theory and automated processes.The paper concludes with the author selfquestion: Would not be better that instead of the Odobleja-Wiener case people would talk about the Odobleja-Wiener cybernetic theory?
Keywords: feedback, cybernetic.

1 Introduction

The rise of a new scientific theory, a significant discovery or a pioneer invention is a process which results from a single person or from more that have created, each person working independent or all together in a team work. The process of generating the new is given by a sequence of stages which may sometimes last tens of years. Often, the filiations of the creative process can be followed, and the chain of the successive influences which led to the rise of a new scientific or technical concept can be established. Within the assessment of the phenomenon represented by the generation of a new science, a series of conjectural elements should be considered, noting, at the same time, the individual contributions of those who played a significant role in the process. A theoretical conceptualisation or pioneer invention is not any different from the technical scientific concerns of the moment. Usually, it is something that "floats in the air", that seemingly announces the new, which is never entirely, 100%, new. (Iancu St., 1996).

At present, it is generally known that Heron of Alexandria (born in the first century AC) designed the steam engine in antiquity and that it became an acute necessity only during the first industrial revolution. Therefore, Giovanni Branca is mentioned in history as the one who suggested, in 1629, that steam should be used as an engine agent of the turbine and T.Savery (1650-1715) as the one who built a usable steam engine, known in the literature as "miner's friend".

The theory of relativity started with Galileo Galilei ideas about inertia and Newton's ideas on the universal gravity, it continued with Jules-Henry Poincare's works, the first to talk about the theory of relativity, with Hendrick Lorentz's theoretical development of J.C.Maxwell's theory and it seemed to end with Einstein's brilliant contribution. Today, the theory of relativity is questioned and new developments are expected to come.

2 The emergence of cybernetics as a science

Born approximately two and a half millennia ago, forgotten until the XIX century and considered the result of a group of specialists that lived between 1921 and 1948, rediscovered in 1948, the word cybernetics received a great deal of attention in the 70s of the XX century. The Greek word kybernetike was used by Plato (427-347 BC) meaning the art of steering a ship, of riding a pair of horses, but also the art of leading people - generally, the art of leading.

In 1834, the French physician and mathematician Andre Marie Ampere introduced in the paper "The Study of the Philosophy of Science" a chart of all the branches of science known until then. In this chart, in the chapter called "Politics", under the column number 83, Ampere placed a new science - kybernetike - meant to deal with "the study of the methods of command and leadership of the society". For each science, Ampere chose a motto made up of Latin verses, cybernetics having as a motto "Et secura cives ut pace fruantur" (And the careless citizens

will enjoy themselves in peace). Indeed, Ampere's amazing vision was confirmed by the subsequent evolution of cybernetics which placed itself on the XX century science orbit through its high gnosiological and praxiologic power, to the benefit of the technical and scientific progress.

The Greek word kybernetike is also used in the religious language, meaning "the science of church organization".

It is difficult to establish in time when exactly cybernetics emerged as a science, since its roots come from a series of old works. Although, rigorously speaking, the philosophical works of the XVII and XVIII centuries in which the living being is compared and sometimes assimilated to the machine can never be considered as belonging to cybernetics; their authors, knowing only automatic machines with a stiff control and not the ones adjusted through reverse connection, presented in their works a material-mechanical conception, totally insufficient in order to explain the cybernetic phenomena.

Until the XIX century the mechanical devices designed to make the adjustment, by maintaining and correcting the movement in order to set and standardize some processes, were made to meet some particular practical requirements and therefore, their operation principle escaped their attention. This explains why until the XIX century no theoretical analysis was made that would highlight the multiplication and improvement possibilities of the mechanisms. The way in which the design of the mechanisms alone evolved until the XIX century is an example in which technique preceded science.

In the XIX century the victory of electricity over steam was proclaimed and adaptive command mechanisms acquired a greater and more diverse use. Consequently, scientists became interested in these devices, using a more refined mathematic machine. The notion of entropy introduced by J.E. Clausius in thermodynamics did not allow its later fertility in the development of information theory to be foreseen. Around 1868, J.C. Maxwell established a mathematic pattern for Watt's regulator which consisted of differential equations that underlined the effect that different system parameters had on its performance, substantiating thus the reverse connection adjustment theory. While some people claim that 1868 is the birth year of cybernetics, others think that Maxwell's paper is nothing else but automation, which has become nowadays, a section of cybernetics, exclusively destined to the artificial automated mechanisms. During the next age, the studies in this direction would intensify and the bases of a mathematical adjustment theory would be constituted in the last decade of the XIX century by using the theory of differential equations.

In the XX century, the mechanic adjustment was replaced with the electromechanical one and subsequently, with the electronic one, noting a significant progress, especially in communications (telephony, radio-technique) but also in fields such as temperature, rotation speed adjustment, etc. Valuable works were designed in Bell laboratories from USA by H.S. Black in 1931 and by H. Nyquist in 1932 in order to develop the telephony and especially the amplifiers. The theoretical studies which addressed automatic regulation issues underlined a fundamental concept that technicians intuited and theorists defined as "feedback". The feedback devices were classified in two categories according to the way in which they led to: the intensification of the effects due to the input signals, regenerating and intensifying the oscillatory processes (positive feedback - for instance the electronic auto-oscillator) or to the reduction of the effect caused by the input signal in order to reduce the deviations, damping them down and achieving process stability (negative feedback - for instance the electronic amplifier that ensures the stability of the application coefficient by reducing the background noise and interior parasites).

The developments in biology allowed the exploration of the adjustment processes in living things: Karl von Voit 1878 - thermal adjustment, J.F.Miescher 1885 - breathing, W.R. Hess 1930 - circulation, I.P.Pavlov (1880-1926) - reflex arch.

In 1932 Jacques Laffitte published the work entitled "Reflections on the science of machines" considered by some people a prediction of N. Wiener's famous work "Cybernetics or command and communication science in beings and machines" and in 1934 Rudolf Carnap treated the language problems. In 1935 M.Kalecki drew the attention on the reverse connection phenomena in economic sciences and Stefan Odobleja published, in 1938/1939, the book "The Consonants Psychology", (Psychologie consonantiste, Librairie Maloine, Paris, vol I 1938, vol II 1939)[1]. In this book, Odobleja thought and enounced the first generalized cybernetic vision. In 1940, W. Schmidt thought of an absolutely general science for the mechanisms with automatic regulation that, according to some researchers, is nothing else but cybernetics, since this would be its birth year.

In 1943 A. Rosenblueth, Norbert Wiener and J. Bigelow published the book "Behaviour, Purpose, and Teleology, Philosophy of Science" in which they presented general cybernetic concepts paying no attention to the mathematical approach and in which the relation between the technical cybernetic processes and those of the

[1]The book "The Consonants Psychology" was reviewed in 1939 in Romania, in the "Modern military spirit" publication and in 1941 in USA in the "Psychological Abstracts" summary magazine.

living organisms was established.

1948 is considered the birth year of cybernetics, since it is the year in which the following papers were published: "Cybernetics or command and communication science in beings and machines" by Norbert Wiener and "The Mathematic Theory of Communication" (which founded the modern theory of information) by Claude Shannon and Warren Weawer. E. Ross Ashby also published "The brain project" in which the homeostatic theory appears.

Wiener did not know the word kybernetike - the Greek equivalent of the art of leading- that had been used by Plato and Ampere for notion cybernetics. Instead, he has used another Greek word - kibernetos, the equivalent of the term pilot, steersman.

It is well known that after the publishing of Norbert Wiener's paper in 1948, cybernetics was applied in technique, economy and sociology. Bio cybernetics, neurocybernetics, economic cybernetics appeared in the 50s and 60s of the XX century .This century came with a cybernetic vision of society and a continuous growth of technical cybernetics. Cybernetic psychology developed in the 70s of the XX century. But, still now, it still encounters obstacles in explaining the psychological level.

The movement of ideas which led to the conversion of cybernetics into a scientific subject was significantly influenced by the accumulation of the results of some researched biological, technical, economic and social processes. One of the biggest achievements of scientists in this respect was the system conception, a global conception, often encountered nowadays, and which led to the establishment of another scientific subject - the general theory of the systems.

It is still under discussion whether cybernetics is an idea, a point of view, a way of thinking or a true science, since it is very difficult to make a distinction between the general theory of the systems and cybernetics because, generally they represent the same field and specifically, cybernetics refers to the structures of the loops with reaction or of the reverse connection (feedback) in a system and to the properties determined by connections. The discussions over the definition of cybernetics will continue but what is clear and amazing about the human way of thinking is the cybernetic concept, the role of the reverse connections in all aspects of reality, their use in technique.

Cybernetics on the whole is considered a science of leadership, the science which is interested in the mathematical study of connections, commands and control within the technical systems and living organisms from the point of view of their formal analogies with a view to design and build automated electronic machines and devices able to perform different operations or operation sequences. Cybernetics could be defined as a synthetic science interested in the mathematical study of the operation of systems characterized by commands and adjustments no matter if they are natural, social or technical systems or as a science of the general evolution and balance laws or as the art of making the action efficient.

The study made by cybernetics on different systems is an abstract study of their formal analogies and not of their constitutive elements or specific functions. Norbert Wiener defines cybernetics' scope stating that it refers to "the entire field of the command and communication theory in machines or animals". Raimond Ruyer insisted, in 1954, on the informative aspect of cybernetics, defining it as "the science of the information machines either natural - the organic ones - or artificial machines" (Ruyer Raymond, 1962).

Norbert Wiener said that "when I lead another person's actions I communicate to him/her a message and even though this message is imperative in nature, the communication technique is not different from the one of transmitting a fact. Moreover, for a command to be efficient I have to know all the messages that came from that person, messages that can announce me that the order is understood and that it had been executed (Wiener Norbert, 1952)". Cybernetics became, thus, the science of information; it is also interested in the general study of the signals or signal systems, always aiming at providing some elements for the study of their transmission and always taking into account the structures of the communication means of these messages (the study of the networks). The two notions - command and communication - mix closely, the information becoming valuable only if it allows acting.

3 Romanian forerunners of cybernetics

It is a great satisfaction to acknowledge that highly valuable cybernetic ideas for the Romanian way of thinking, culture and science arose in the mind of four important representatives: Spiru Haret, Daniel Danielopolu, Stefan Odobleja and Paul Postelnicu.

3.1 Spiru Haret (1851-1912)

Romanian philosopher, mathematician and sociologist published in 1910 in Paris his masterpiece "Sociale Mecanique" (Spiru Haret, 1969)", the first big Romanian work which addressed social-economic issues with the help of the mathematical patterns and which analysed these processes according to the system. The book "Sociale Mecanique" is not an attempt to apply the entire body of axioms and laws belonging to the rational mechanics to the social movement, Spiru Haret trying to adapt only the device of the rational mechanics to the analysis of the social and economic phenomena evolution, continuously revealing the specific nature of these phenomena and the fact that they must not be confused with the mechanical phenomena.

Spiru Haret used in his paper several mathematical patterns that refer to some demographic problems, to the role of science and technique in the development of the society, etc. and defined society using a cybernetic model which includes both the recreation idea and that of data collection and processing in order to carry out command and control processes. Obvious cybernetic interpretations are given by Spiru Haret in the examination process of the social balance, in the treatment of the natural population growth effect on the social life, interpretations which led to the conclusion that an unwanted state of social rest could not be achieved in civilized societies and that if this was possible it would be unstable.

Spiru Haret's believes about the periodicity of social phenomena are very interesting from the point of view of the current general theory of the cybernetic systems. Making reference to a society which has a certain evolution and which finds itself under the influence of a trend or group of new necessities that will impose a new direction, Spiru Haret shows that a swinging movement will take place in that society and that the moment it reveals itself through what is known as "action and reaction" it will push the societies in one direction or another when the relevant ideas intervene or the pressure of the new necessities exercise until the movement melts down and disappears in the general movement of the system" (Spiru Haret, 1969).

Spiru Haret's work underlines through its numerous examples the conception of this great forerunner of our national system science and culture.

3.2 Daniel Danielopolu (1884-1955)

Romanian physiologist reached cybernetic conceptions following the path of biology and medicine and dedicated his work to the experimental analysis and logic schematization of the nervous, endocrine and immunity system.

Elements of his cybernetic thinking began to appear in 1923, determining a bio-systemic, cybernetic vision of the human system. Without making a clear distinction between the informational and the non-informational systems, Danielopolu intuitively felt it and showed a predilection for the former. Studying for over 50 years biosystems that receive, send process and generate information, he highlighted and deduced the existence of some mechanisms, some operation principles or some operational laws that coherently integrate in bio cybernetics. In the absence of the formal apparatus that we have nowadays and without using the modelling technique, Danielopolu made in his works synthetic interpretations and graphic and logic schematizations which constitute an exceptional scientific prevision of bio cybernetics.

All Daniel Danielopolu's works regarding the neurone-endocrine and immunity systems were based on a systemic interpretation. In the 20s of the XX century he elaborated a theoretical model applicable to the way of functioning of the neurone-endocrine and immunity systems. In 1928, Danielopolu formulates the following three laws: the law of the amphomechanism, the law of the predominance and the law of the circular mechanism which lay at the basis of the interstimulative antagonism. The analysis of these laws highlights the fact that they imply the existence of some negative reverse connections in the neurovegetative system. Without calling it negative reverse connection, Danielopolu identified this type of reaction in order to explain the operation mode of other bio systems as well. In some works, published between 1923 and 1932, he intuited the concept of positive reverse connection and its importance in physiopathology.

Daniel Danielopolu was an international forerunner of bio cybernetics, a forerunner of the biological systems theory and of cybernetic medicine.

3.3 Paul Postelnicu (1917-1983)

An electro mechanic engineer (graduate of the Polytechnic School from Bucharest in 1941) with philosophical inclinations, employee of the Romanian Telephone Society, managed in 1945 to independently present a cybernetic vision on life. The cybernetic loop - according to Paul Postelnicu - was a characteristic of the matter in general.

In 1944, Paul Postelnicu drew up an article "The Theory of the Vicious Complex" that he sent to "The Scientific Magazine - V. Adamachi" from Iasi (3 pages) to be published. In the absence of a response in real time he rewrote and developed a version of 6 pages entitled "The Hypothesis of the Vicious Complex" that he sent for publication in 1945 to the "Nature" magazine from Bucharest. The article, with its two versions, circulated in the typed form among its colleagues and it represented the foundation of a paper presented by Paul Postelnicu within the meeting of the "Friends of natural sciences" society of 24 February 1945. The article appeared only later, in 1968, in The Telecommunications magazine. (Paul Postelnicu, 1968)

Paul Postelnicu's basic idea relies on the concept of a vicious circle that he called "vicious cycle", maybe in order to avoid the confusion with the vicious circle from the elementary logics. He defines the vicious cycle as a causal chain made up of a, b, c....m, n,.... phenomena which have the property that a determines b, b determines c and so forth. These causal chains can be exemplified with systems from physics such as: engines, reaction electronic tubes, etc. Any system that includes a vicious cycle is called by Paul Postelnicu "a vicious complex".

Paul Postelnicu determined the properties of the vicious cycles after he had made a detailed analysis of the amplifying triode with feedback loop underlying the relation between the circuits of the grid and of the plate, where one's oscillations determined the others and the other way around and finally, he concluded that "the viciousness" of the triode characterized the biological phenomena as well.

On the basis of an identified property of the vicious cycle, that is that "any progressive vicious cycle includes the resonance condition", Paul Postelnicu stated that hazard made so that a little material part would acquire an organization that would essentially differentiate it from the inert matter. Due to the resonance condition the vicious cycle became progressive and evolved through impulses that came both from the hazard and the environment. Thus, he admitted that the appearance of the molecule with vicious cycle properties could be explained.

Paul Postelnicu asserted that matter became alive when in its organisation a system of reverse connections is established (bio cybernetic conception). "Viciousness is an essential property of the matter and even the evolution of the universe itself could be explained by the hypothesis of the vicious complex. In this case, matter and life would not be but two levels of vicious organisation: the matter as a vicious organisation of energy, life as a vicious organisation of matter and implicitly of energy"(Postelnicu Paul, 1968). Starting form the existence of the "viciousness" in physical devices created by man, Paul Postelnicu extended it to the biological, social and economic phenomena and then he generalized it to the phenomena that go beyond the usual human experience. Under this form, the viciousness was admitted as a constitutive element of a general theory that would allow the construction of a cybernetic model of the universe, eventually in the sense of the distinctions operated by the general theory of the systems.

3.4 Ştefan Odobleja (1902 -1978)

Is one of the great thinkers and creators of the XX century whose originality and clear-sightedness overcame the traditional scientific knowledge patterns and models.

The life and his selected works

Born on 13 October 1902 in Izvorul Anestilor -Mehedinti, Odobleja wins in 1922 a scholarship at the Faculty of medicine and becomes scholar of the Military Medical Institute. During his studies he makes researches in neurology, psychology and methodology of knowledge and logics. In 1928 he becomes PhD in medicine and surgery maintaining a thesis on car accidents and on 1 May 1929 he publishes the study entitled "Method of thoracic transonance" in the "Medical Therapeutic Bulletin" in which he formulates the so-called reversibility law.

Between 1928 and 1935 he publishes in the specialized magazines of that time a series of works regarding the study of the organism using the method of listening to the noises of the human body, works that he gathers in a volume called "La Phonoscopie, nouvelle methode d'exploration clinique". The volume published by Gaston Doin Publishing House - Paris was awarded by the Romanian Academy with the "General Physician Dr. Papiu Alexandru" prize, which was granted every 2-4 years to the most praiseworthy works written by military physicians;

In 1936 Odobleja publishes the work entitled "Phonoscopy[2] and the clinical semiotics" and in 1937 he participates in the IXth International Congress of Military Medicine with a paper entitled "Demonstration de phonoscopie" and on this occasion he disseminates a prospectus in French announcing the appearance of the work "The Consonantist Psychology".

[2]**Phonoscopy** - the photographic recording of intrathorax noises (in the heart or lungs)

"The Consonantist Psychology"[3] was elaborated between 1929 and 1937, the author starting from the idea of the possibility of studying a phenomenon (such as the moods of the animal organism) through other phenomena (such as the phonic ones). The book was published in two volumes comprising 880 pages (Psychologie consonantiste, Librairie Maloine, Paris, vol I 1938, vol II 1939. The paper was printed in Lugoj). Soon after its publishing, the book was reviewed in 1939 in Romania in the "Modern military spirit" publication and in 1941 in USA in the "Psychological Abstracts" summary magazine.

During the war, Odobleja was chief physician on a military ambulance and after the war, he expressed his intention of working in research but, completely misunderstood, he was forced in 1946 to retire from the army. After his retirement, he did not manage to find a job appropriate to his education and wishes and he continued to live on a modest military pension.

Ion Oancea - Stroe, Odobleja's friend, tells that "I met the author in 1949-1952 in his parents' house from Valea Izvorului. During our short introduction he never told me that he wrote such important works. Only later, around 1964, when he changed his residence in Turnu Severin did he carry with him the volumes of "The Consonantist Psychology", a copy of which he had given to the "I.G.Bibicescu" Municipal Library. After 1968 he participated in "Al. Vlahuţă" Literary circle, where he read some passages from his book" (Oancea - Stroe, 1972).

Odobleja meditated on the issues discussed in his book "The Consonantist Psychology" and reached the conclusion that he was the possessor of an original conception on logics, different both from the traditional Aristotle one and the modern mathematic one and started to write a logic paper that he abandoned after 1973. The totality of Ştefan Odobleja's manuscripts exceeds 50,000 files.

In 1975 - during the Third Cybernetics International Congress that took place in Bucharest - Odobleja presented the paper "Cybernetics and the consonantist psychology" and in 1977 the text of his paper appeared in the volumes published by the Congress (Springer Verlag and The Technical Publishing House). In 1978 he sent to the IVth Cybernetics and Systems International Congress that was taking place in Amsterdam, the paper entitled "Diversity and unity in cybernetics".

Ştefan Odobleja's cybernetic conception

"The Consonantist Psychology" is not a cybernetic or purely psychological paper; it is a work of thinking, a philosophy of mental processes and of science, the author searching several general laws: laws of movement, balance and unbalance, energetic processes, reversibility and irreversibility, etc. that would apply to all domains, all inert or living natural sciences, to psychology, socio-economic life. Among the fundamental laws that, according to Odobleja, govern the physical and the psychic, the reversibility law or as we might call it nowadays, the reverse connection law, or feedback law is worth mentioning. The author of "The Consonantist Psychology" treats in an interdisciplinary, integrating vision, issues related to psycho-neurology, psycho-physiology, psycho-pathology, psycho-therapy and therapeutics, on the one hand, and issues related to the physical-chemical, mathematical, biological, sociological economic, philosophical, etc. sciences on the other hand and also issues related to their specific laws, that are in close interdependence and interconditioning, focusing especially on the connections from the physical-psychological-philosophical levels.

Ştefan Odobleja thought and enounced the first generalized cybernetic vision. In order to appreciate the importance of Ştefan Odobleja's cybernetic conception we must see two possibilities: either Odobleja knew the importance of different cases of reverse connections belonging to diverse fields of science and technique and made a generalization or he obtained the general vision not so much through a generalization, which maybe was not sufficiently prepared in 1938, but through a new approach of the phenomena belonging to the death or living nature. Odobleja made references to the important role played by the improvement of the observability in knowledge and by the development of stability, which nowadays in the cybernetic language means the role of increasing the information quantity in the improvement of systems' stability.

Obviously, Ştefan Odobleja's starting point in the creation of his cybernetic vision was psychology. This also justifies the name of his work. This was an original idea unlike the attempts of that time to substantiate the sciences and which were relying either upon mathematical logics or upon linguistics. To the extent in which any science relies upon the creative activity of the human mind it is at the same time a psychic activity but not a regular one, one that is elevated and logical-psychical.

[3]**Consonantist** - harmonious. Ştefan Odobleja's consonantism is the theory of man's psychic activity. All psychic phenomena are consonance, that is, the psychic is also physical, exiting however a physical-psychic consonance- where the physical is understood in its ordinary meaning and the psychic is another type of physic. Since consonance means harmony, the consonantist psychology becomes, according to the author, a logic of harmony or in other words, a science of organisation and self-adjustment

The substantiation of the sciences on the basis of psychology made Odobleja, through his conception, the forerunner of cybernetics, of a technique of thinking in general and not only of the mechanisation of the mathematical calculus. Determining the relations of consonance among sciences, multiple (spatial, physical, logical and psychological) consonances, Odobleja postulates the possibility of mechanising the thinking - conceived as a creative act - and suggests a thinking machine.

Odobleja was not only a forerunner of cybernetics but he also had an original vision in this field, regarding both the mechanization and automation of the constituted, repetitive thinking and the creative, authentic thinking. Odobleja imagines not only thinking machines but also machines that could create, invent, philosophize. Numerous manuscripts on logics refer exactly to this possibility which presumes not only a special psychology, a cybernetic one but also a special logic that he called resonance logics.

Odobleja, through his consonants psychology conceived a general cybernetics with a much larger sphere than Wiener's cybernetics and managed to create the basic paradigm of the cybernetic thinking which requires to focus on the study of the adjustment processes and on the analysis of the behaviour of "any" system by correlating the direct (command-execution) connection with the reverse (execution-command) connection. This circuit was called by Odobleja "vicious circle" and by Wiener feedback (reverse connection principle) (Wiener Norbert, 1966).

Ştefan Odobleja wrote that "There is reversibility and reciprocity of interests between individual and society. Vicious good or bad circles are often established between the individual and society", and ten years later, Norbert Wiener showed that "Communicating with the outer world means receiving messages from it and sending it messages. On the one hand, it means observing, experimenting and learning and on the other hand it means exercising our influence on it so that our actions to become subordinate to a purpose and efficiency... Life is a continuous interaction between individual and its environment and not a way of being eternal." (Wiener Norbert, 1972).

While Ştefan Odobleja's starting point in the creation of his cybernetic vision was psychology, from the cybernetic structure of organisms' biological and psychological processes towards the adjustment and control technical systems but also towards economy and society, Wiener, who had worked in a technological institute being concerned with the use of the mathematical methods in electronics, started from the cybernetic structure of the adjustment and control technical systems towards the biological systems. The difference between "vicious circle" and "reverse connection principle" is purely linguistic, since both concepts have the same semantic content and operational meaning, that is, the influence that the effect exercises, in its turn, on the cause which caused it.

Odobleja sustained that "Reversibility is a vicious circle between the cause and its effect. It is an oscillation between two states which alternatively induce themselves one another. Reciprocity of actions ... Besides the acyclic causality there is a cyclic one, in a vicious circle. The same phenomenon is then one at a time effect and cause". Wiener stated that the cause-effect influence can be characterised in three types of phenomena: reduction / annulment of system's deviation from its state of balance and its coming back to the initial state - homeostasis-negative feedback; the amplification of the distance between the initial state (of weak organisation) and the final state (of better organisation) - development - positive feedback and a larger distance between the initial state of the system of good organisation and the final state of disorganisation - involution - positive feedback. Odobleja made no clear distinction between the positive retroaction and the negative one, although it results from the book that he intuited their different role.

Odobleja's vision also had a series of limitations inherent to the age he lived in. Ştefan Odobleja operated with the diode "Substance-Energy", Norbert Wiener with the triode "Substance-Energy-Information", the latter concept having a special value. Firstly, the lack of the notion of information, not the information found in the statistic theory of communications but in its widest sense, represented a deficiency of Ştefan Odobleja's cybernetic vision. Secondly, Odobleja had no contact with electronics or automation and therefore he did not know the principles of the positive and negative feedback discovered in electronics before the period in which he elaborated his works.

Before any formalism, it is necessary to clarify the essential processes to which the mathematical treatment would apply and Odobleja, not having all the essential cybernetic notions did not use any mathematical formulation in his work and, therefore, the Consonantist Psychology did not have the scientific impact of Wiener's paper.

An analysis of great perspicacity on ethics, present in Doctor Ştefan Odobleja's work "The Consonantist Psychology" reveals the fact that the author presented different points of view in order to define and implicitly understand the ethics which was defined as: "the science of good and happiness, the science of social balance, the science of morality and immorality, of rights and duties, of vices and virtues, the science of agreement and of consonance between the individuals' interests or between the individual's interests and the interests of society, the science of harmony and balance between the self and the society, the science of the "confrontations " between individuals and the ratio of avoiding them, the science of prudence and carefulness, the science of approximating its own (actual or future, near or remote) weakness. Ştefan Odobleja, never famous or appreciated at his true value in his own

country, presents us in a maximum synthesis the moral possibilities. The most eloquent and the deepest definition that Odobleja gives to ethics must be pointed out. "Ethics is the science of the prophylaxis and therapeutics of the evil". This definition can be considered the most appropriate one since ethics is defined through its mission.

Globally acknowledging Odobleja's work

After 1948 some Marxist philosophers showed towards cybernetics an obvious negative attitude and, that is why, cybernetics had no citizenship right in the scientific circles of the socialist world. Cybernetics was considered in this world, for approximately 15 years, a "pseudo-science, strictly mechanicals, reactionary and useless". After 1964, a period of relaxation begins in Romania and Norbert Wiener's paper "Cybernetics" is published in 1966 at the Scientific Publishing House from Bucharest and in 1967 Jacques Guillaumaud's paper entitled "Cybernetics and dialectic materialism" appears with a preface signed by Paul Langevin. In this book is presented a short history of the cybernetic ideas with no reference to Odobleja.

Odobleja's work attracted again the attention of the Romanian public after many years due to I. Stroe - Oancea and to V. Pârvulescu who published in the "Future" newspaper from Mehedinti, on 12 May 1972 the article "Severinean Micro encyclopaedia".

Dr. Al Olaru presented in 1973 at Bucharest within a conference regarding the history of medicine, a paper on "The Consonantist Psychology" and on 11 May 1974, Constantin Bălăceanu Stolnici published in "Flacara (Flame)" magazine the article entitled "The Romanian Ştefan Odobleja, a pioneer of cybernetics".

Odobleja led a solitary life rediscovering himself, once with the appearance of the first books on cybernetics in Romania and, especially, after the publishing, in 1966, of Norbert Wiener's work. Convinced of the value of his ideas, Ştefan Odobleja addressed after 1972 to the Romanian Ministry of Foreign Affaires, asking for support in the elucidation of a Romanian priority and for the translation of his book in the Romanian language. Following his request and on the background of the discussions had in mass-media, the first debates took place and the first studies on Odobleja's work began to be published. In 1974, Professor Dr. Constantin Bălăceanu Stolnici states that Ştefan Odobleja is an important forerunner of cybernetics (Constantin Bălăceanu Stolnici, 1974). In 1975 V. Săhleanu publishes the article "Daniel Danielopolu and Ştefan Odobleja forerunners of cybernetics" (Săhleanu V, 1975) and Professor Eugen Niculescu Mizil in "Studies of social and economic cybernetics" notes the strong cybernetic character of the work "The Consonantist Psychology", in which Ştefan Odobleja enounced the law of the feedback circuits that he generally emphasized in all the processes and phenomena from nature and society.

After a detailed analysis of Ştefan Odobleja's work, the Section of Medical Sciences of the Romanian Academy through its Interdisciplinary Research Group reaches the conclusion that Ştefan Odobleja's capacity of forerunner of cybernetics could be taken into account at the beginning of 1975, considering the ideas of "reversibility" from "The Consonantist Psychology", similar to those that led to the appearance of cybernetics but there cannot be made a connection of filiations between the ideas from "The Consonantist Psychology" and those from Norbert Wiener's "Cybernetics" published in 1948.

On 28 September 1977, the academician Mihai Drăgănescu delivers a lecture at the County Library from Mehedinţi, Drobeta Turnu Severin, on "Ştefan Odobleja - the relation between man and machine", and in 1978 Ştefan Odobleja publishes at Craiova, at the Romanian Writing Publishing House the work "The consonantist psychology and cybernetics", in which he states his own ideas regarding his priorities in cybernetics.

All the studies and papers written during the period in which "The Consonantist Psychology" was analysed were gathered in the volume "Romanian forerunners of cybernetics" published in 1979. This volume contributes to the recognition at the Romanian academic level of Ştefan Odobleja's merits as forerunner of cybernetics and in 1981 the Romanian Academy dedicates to him the collection of studies "Odobleja between Ampere and Wiener" in which Odobleja is placed as a forerunner between the two internationally known savants.

In 1982 the Scientific and Encyclopaedic Publishing House from Bucharest published for the first time in Romanian the work "The Consonantist Psychology" with an introductory study signed by the academician Mihai Drăgănescu and professor Pantelimon Golu, and in 1984 Odobleja's paper "Introduction to resonance logics" appears in Craiova, The Romanian Writing Publishing House, edition revised by the academician Alexandru Surdu, preface signed by Constantin Noica (a selection of the first 4,000 leaves of the manuscript "Resonance Logics", which includes over 15,000 leaves).

In a series of studies meant to examine Ştefan Odobleja's role and position in the development of cybernetics the academician Mihai Drăgănescu concluded that: "Ştefan Odobleja cannot be considered the founder of cybernetics, this being Norbert Wiener's merit, but Ştefan Odobleja not only is he a forerunner of cybernetics but he also has world priority of the idea of a generalized cybernetics being the first to consider the closed loop phenomenon, therefore with reverse connection, as a universal law. According to our knowledge no one else before him had

such a vision of the general role of feedback (reversibility law) in as many fields as possible, generally, in all fields. In this way, he delimits himself from all the particular instances in which reverse connections were emphasized." (Drăgănescu M, 1982).

On 13 November 1990 Ştefan Odobleja is chosen post mortem member of the Romanian Academy.

Ştefan Odobleja's manuscripts wait to be exploited. Constantin Noica in the preface to "Introduction to resonance logics" said that the publication of Odobleja's works "could contribute to the renewal of the perspectives and points of view of today's scientific culture, a culture within which Odobleja's original ideas cannot be ignored. If he could not make the most of them until the end, we are convinced that, in the Romanian culture at least, through their novelty and boldness they will awake great creative vocations that we need to qualify in world's culture".

The international recognition of Odobleja's work

"The Consonantist Psychology" was reviewed in 1941 in USA in "Psychological Abstracts" summary magazine.

In August 1978 the IVth International Congress of cybernetics and systems took place in Amsterdam but because Ştefan Odobleja was ill he could not participate in person but he sent the paper entitled "Diversity and Unity in Cybernetics" that was presented at the congress by a Romanian representative and in the same year the text of the paper appeared in the volumes published by the Congress (Springer Verlag).

B. H. Rudall from the University of Wales, who presided the session where the paper was presented, said: "Dr. Odobleja's work was very well received... ... Dr. Odobleja's precedent paper was considered very interesting and it was highly appreciated ("The Consonantist Psychology"), but of course there were no formal discussions on any pretences, and although unofficial discussion were carried out after our meetings, I do not know yet enough about dr. Odobleja's contributions in cybernetics for me to comment on his work ".

Thus, at the IVth International Congress of cybernetics and systems dr. Ştefan Odobleja's case as a forerunner of cybernetics was raised before the entire scientific community and consequently, his international life began in 1978, once with Odobleja's death.

4 Summary and Conclusions

1. Cybernetics or the theory command and communication in beings and machines is the creation of a group of experts belonging to different fields who between 1920 and 1948 observed that a series of problems related to the control of the machines and organisms have in common certain organisation mechanisms and laid the foundations of a new subject evolving around two concepts: feedback - reverse connection principle (feedback circuit) and information.

2. Through his paper "The Consonantist Psychology", Ştefan Odobleja proved that he had genius, that he deserved to appear in the universal science besides Ampere and Norbert Wiener in the establishment of the cybernetic concept and way of thinking. Ampere anticipated cybernetics as a science, Ştefan Odobleja elaborated the central ideas of cybernetics and the cybernetic way of thinking, N. Wiener regained what, in other conditions, Ştefan Odobleja has discovered and managed to build cybernetics as a science through a complete mathematical analysis of the feedback theory and automated processes.

3. The second half of our century is marked by cybernetic concepts because of Norbert Wiener, but behind him are many scientists as in any other field of knowledge. Among these, Ştefan Odobleja is worth mentioning, as he is the one of its most valuable moments. Ştefan Odobleja's biggest merit is the one of having discovered the general character of the feedback and of having tried to emphasize it in the most diverse range of processes and phenomena.

4. Odobleja made the first step in intuiting the cybernetic science but it was an important one and when it became internationally known, the Odobleja - Wiener case was born. Maxwell remains the founder of electromagnetism but history mentions Faraday as the one who intuited the electromagnetic waves before him. Why should not history mention the fact that Odobleja intuited a generalized cybernetics before Wiener. Wouldn't it be better that instead of the Odobleja - Wiener case people would talk about the Odobleja-Wiener cybernetic theory?

References

[1] Iancu Şt., "Ştefan Odobleja," *Journal "Academica"* , nr. 6,7,8/(66,67,68), aprilie, mai, iunie, p. 62,1996.

[2] Bălăceanu Stolnici C., "Un pionier al ciberneticii, românul Ştefan Odobleja," *Journal "Flacăra"* , 11 May 1974.

[3] Drăgănescu Mihai, "Conceptele cibernetice ale lui Ştefan Odobleja," *Introductory study to the volume: Ştefan Odobleja, The Consonantist Psychology* , The Academy Publishing House, 1982.

[4] Oancea Stroe I., Pârvănescu V., "Microenciclopedie severineană," *Journal "Viitorul"*, Mehedinţi, 12 May 1972.

[5] Postelnicu P., "Ipoteza complexului vicios," *Journal "Telecomunicaţii"*, 1968, vol.12, no. 12.

[6] Ruyer Raymond, "La cybernetique et l'origine de l'information," Publishing House Flammarion, Paris, 1962.

[7] Săhleanu V., "D. Danielopolu şi Ştefan Odobleja precursori în cibernetică," *Journal "Tribuna"*, 30 January 1975.

[8] Spiru Haret, "Mecanica socială," The Scientific Publishing House, Bucharest, 1969.

[9] Norbert Wiener, "The Human Use of Human Beings," French Edition: Cybernetique et Societe, Union Generale d'Edition, Paris, 1952.

[10] Wiener Norbert, "Cybernetics," 1966, The Scientific Publishing House, Bucharest.

[11] Wiener Norbert, "Sunt matematician," The Political Publishing House, Bucharest, 1972.

Ştefan Iancu
Information Science and Technology Section of the Romanian Academy,
Str. Sibiu No.4, Sector 6, Bucharest, Romania
E-mail: stiancu@acad.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 92-107

Colony of robots: New Challenge

Workshop invited key lecture

Gastón Lefranc

Abstract: The evolution on Robotics has in a cross way of application. For one side is the applications to manufacturing, other one is application to medicine, another one in space exploration and it is starting home applications. It is very popular to have contest of robots for students, motivating very well to student, supported by universities, achieving good image for the institutions.
One way for mobile robots is Nomad, a nice application for having new knowledge in the space, but the inversion it is very expense and complex. If it has problem or fail, all the work will stop. Instead of that, if you use a community of robots, working like a society of insect, it is possible to have simpler mobile robots to have specific tasks, less expensive, more reliable to reach the same aims.
In this presentation is focusing in colony of robots. This implies to merge several disciplines based on models of communities, to have control of a society of robots working together in a collaborative and cooperative way in non structured environments.
Keywords: Multi-robots, Colony of robots, Multiagents Systems

1 Introduction

There exists a Nomad robot used to do all mission in Mars. This mobile robot is a nice application for having new knowledge in the space, but the inversion it is very expense and very complex in design: It requires several capabilities to operate in many environments. The cost was over 1.6 million dollars. If it has problem or fail, all the work will stop. An easier and cheaper way, is to have a groups of mobile robots working together to accomplish an aim that no simple robot can do alone.

An ideal application for groups of heterogeneous robots working together, like a society of insect, can accomplish the same mission that one robot. Using simpler mobile robots doing specific task, is less expensive, more reliable and it can reach the same aims of one robots. Some examples of applications are in manufacturing, medicine, space exploration and home.

The nature of work environments requires the robotic systems be fully autonomously in achieving human supplied goals. One approach to designing these autonomous systems is to develop a single robot that can accomplish particular goals in a given environment. The complexity of many environments or works may require a mixture of robotic capabilities that is too extensive to design into a single robot. Additionally, time constraints may require the use of multiple robots working simultaneously on different aspects of the mission in order to successfully accomplish the objective. In cases, it may be easier and cheaper to design cooperative teams of robots to perform the same tasks than it would be to use a single robot. Then, it is possible to build teams of heterogeneous robots that can work together to accomplish a mission, where each robot has different architecture performing different task in a collaborative manner.

Any of this group of robots needs reliable communication among them, in such way that the robots will be able to accomplish their mission even when no robot failures occur. The multi robot system required some knowledge of capabilities of its team-mates, before the start of the mission.

The team of robots can be model observing the natural behaviour of insects. They form colonies with individuals that perform different roles in function of the needing of the community. Using this model, it is possible to have colony of robots with some robots in charge of some responsibilities to work with others in a cooperative way to do same tasks, in a collaborative way, to communicate each other to be more efficient or to take decisions in a collective way, etc. in the same form as natural insects. This colony has to have "nest" where the some robot assigns to others what to do, and other robot that receive orders from human and communicate him the results.

It is necessary to formulate, describe, decompose, and allocate problems among a group of intelligent agents; to communicate and interact; to act coherently in actions; and to recognize and reconcile conflicts. In this presentation is focusing in colony of robots. This implies to merge several disciplines such as mobile robotics, intelligent agents, ontologisms, semantics, as well as automatic control, models of communities, communication, and others ones, to have control of a society of robots working together in a collaborative and cooperative way in a non structured environments.

2 Colony of robots

Having a colony of robots has many advantages over a single robot in cost, complexity and capabilities. The group of robots has to be reliable and to act, in some cases, simultaneously, each doing different tasks in a cooperative and collaborative work. Those groups of cooperating robots have proven to be successful at many tasks, including those that would be too complex for a single robot to complete the objectives. Furthermore, whereas a single complex robot can be easily crippled by damage to a critical component, the abilities of a colony can degrade gradually as individual agents are disabled. In nature, some of the most successful organisms survive by working in groups.

The multi robot system has to have Robustness, Fault tolerance, Reliability, Flexibility, Adaptively and Coherence. Robustness refers to the ability of a system to gracefully degrade in the presence of partial system failure. Fault tolerance refers to the ability of a system to detect and compensate for partial system failures.

The group of robot requires robustness and fault tolerance in a cooperative architecture emphasizes the need to build cooperative teams that minimize their vulnerability to individual robot outages. The design of the cooperative team has to ensure that critical control behaviours are distributed across as many robots as possible rather than being centralized in one or a few robots. The failure of one robot does not jeopardize the entire mission. It must ensure that each robot should be able to perform some meaningful task, even when all other robots have failed, on orders from a higher level robot to determine the appropriate actions it should employ. The design has to ensure that robots have some means for redistributing tasks among themselves when robots fail. This characteristic of task reallocation is essential for a team to accomplish its mission in a dynamic environment.

Reliability refers to the dependability of a system, and whether it functions properly each time it is utilized. One measure of the reliability of the cooperative robot architecture is its ability to guarantee that the mission will be solved, within certain operating constraints, when applied to any given cooperative robot team.

Flexibility and adaptively refer to the ability of team members to modify their actions as the environment or robot team changes. Ideally, the cooperative team should be responsive to changes in individual robot skills and performance as well as dynamic environmental changes. In addition, the team should not rely on a pre-specified group composition in order to achieve its mission.

The robot team should therefore be flexible in its action selections, opportunistically adapting to environmental changes that eliminate the need for certain tasks, or activating other tasks that a new environmental state requires. The aim is to have these teams perform acceptably the very first time they are grouped together, without requiring any robot to have prior knowledge of the abilities of other team members. [27]

Coherence refers to how well the team performs as a whole, and whether the actions of individual agents combine toward some unifying goal. The coherence is measured along some dimension of evaluation, such as the quality of the solution or the efficiency of the solution [4]. Efficiency considerations are particularly important in teams of heterogeneous robots whose capabilities overlap, since different robots are often able to perform the same task, but with quite different performance characteristics. To obtain a highly efficient team, the control architecture should ensure that robots select tasks such that the overall mission performance is as close to optimal as possible.

The key issues for a colony of robots are: Fault tolerant cooperative control; Distributed control; Adaptive action selection, the importance of robot awareness, Inter robot communication, and improving efficiency through learning, Action recognition and Local versus global control of the individual robot Previous research in fully distributed heterogeneous mobile robot cooperation includes: [26] who proposes a three layered control architecture that includes a planner level, a control level, and a functional level; [6], who describes an architecture that includes a task planner, a task locator, a motion planner, and an execution monitor; [3] who describes an architecture that utilizes a negotiation framework to allow robots to recruit help when needed, [7], who uses a hierarchical division of authority to perform cooperative fire-fighting. However, these approaches deal primarily with the task selection problem and largely ignore the issues so difficult for physical robot teams, such as robot failure, communication noise, and dynamic environments. Since these earlier approaches achieve efficiency at the expense of robustness and adaptively, [27] emphasizes the need for fault tolerant and adaptive cooperative control as a principal characteristic of the cooperative control architecture.

3 Models for Colony of robots.

It is necessary have a model of the community, normally taken from the low level intelligence, such as ants, bees or other insects. With one of those models, the robots used are limited in capabilities, such as limitations in computation, actuation and sensing capabilities.

Many colonies of robots have successfully demonstrated cooperative actions, most notably in [25] and [14]. But while much of the existing research has centred on highly-specialized, expensive robots, others create a colony with simple, small and inexpensive robots. [11] propose the developing a robot colony, with intelligent distributed behaviours, with the goals: low-cost robots, homogeneous architecture and distributed algorithms.

Multiple agents can often accomplish tasks faster, more efficiently, and more robustly than a single agent could. Groups of cooperating robots have proven to be successful at many tasks, including those that would be too complex for a single robot to complete the aims. Multi-agent systems have been applied to such diverse tasks as complex structure assembly [15], large-object manipulation [40], [37], distributed localization and mapping [13], multi-robot coverage [9], and target tracking [16]. Furthermore, whereas a single complex robot can be easily crippled by damage to a critical component, the abilities of a colony can degrade gradually as individual agents are disabled. In nature, some of the most successful organisms survive by working in groups.

4 Behaviour of Colony of robots.

The behaviour approach to autonomous robot control is based on the observations of animal behavior, particularly the lower animals, obtaining results without the need for a complex, human-level architecture. Since there are so many varieties of social behavior in the animal kingdom, the classification proposed by Tinbergen [41] is of particular interest for current robotics research in cooperative systems, as it parallels two possible approaches to cooperating mobile robot development. According to Tinbergen, animal societies can be grouped into two broad categories: those that differentiate, and those that integrate. Societies that differentiate are realized in a dramatic way in the social insect colonies [42].

These colonies arise due to an innate differentiation of blood relatives that creates a strict division of work and a system of social interactions among the members. Members are formed within the group according to the needs of the society. In this case, the individual exists for the good of the society, and is totally dependent upon the society for its existence. As a group, accomplishments are made that are impossible to achieve except as a whole. On the other hand, societies that integrate depend upon the attraction of individual, independent animals to each other. Such groups consist of individuals of the same species that "come together" by integrating ways of behavior [31]. These individuals are driven by a selfish motivation which leads them to seek group life because it is in their own best interests. Interesting examples of this type of society are wolves and the breeding colonies of many species of birds, in which hundreds or even thousands of birds congregate to find nesting partners. Such birds do not come together due to any blood relationship; instead, the individuals forming this type of society thrive on the support provided by the group. Rather than the individual existing for the good of the society, we find that the society exists for the good of the individual.

There are to approach to model cooperative autonomous mobile robots, with a behavior similar to classifications of animal societies discussed above. The first approach involves the study of emergent cooperation in colonies, or swarms, of robots, an approach comparable to differentiating animal societies. This approach emphasizes the use of large numbers of identical robots that individually have very little capability, but when combined with others can generate seemingly intelligent cooperative behavior.

Cooperation is achieved as a side-effect of the individual robot behaviors. A second approach parallels the integrative societies in the animal kingdom, aims to achieve higher-level, "intentional" cooperation amongs robots. In this case, individual robots that have a higher degree of "intelligence" and capabilities are combined to achieve purposeful cooperation. The goal is to use robots that can accomplish meaningful tasks individually, and yet can be combined with other robots with additional skills to complement one another in solving tasks that no single robot can perform alone. To be purely analogous to the integrative animal societies, robots in this type of cooperation would have individual, selfish, motivations which lead them to seek cooperation [McFarland, 1991]. Such cooperation would be sought because it is in the best interests of each robot to do so to achieve its mission. Of course, the possession of a selfish motivation to cooperate does not necessarily imply consciousness on the part of the robot. It is doubtful that we would attribute consciousness to all the integrative societies in the animal kingdom; thus, some mechanism must exist for achieving this cooperation without the need for higher-level cognition. The type of approach one should use for the cooperative robot solution is dependent upon the applications envisioned for the robot team. The differentiating cooperation approach is useful for tasks requiring numerous repetitions of the same activity over a relatively large area, relative to the robot size, such as waxing a floor, agricultural harvesting, cleaning barnacles off of ships, collecting rock samples on a distant planet, and so forth. Such applications would require the availability of an appropriate number of robots to effectively cover the work area while continuing to maintain the critical distance separation. On the other hand, the intentional cooperation approach would be

required in applications requiring several distinct tasks to be performed, perhaps in synchrony with one another. Throwing more robots at such problems would be useless, since the individual tasks to be performed cannot be broken into smaller, independent subtasks. Examples of this type of application include automated manufacturing, industrial-household maintenance, search and rescue, and security, surveillance, or reconnaissance tasks.

5 Multirobots Systems.

In a multirobots systems have been identified seven primary research topics, when the colony has distributed mobile robot systems. This issue are the following: biological inspirations, communication, architectures, localization/mapping/exploration, object transport and manipulation, motion coordination, and reconfigurable robots.

5.1 Biological Inspirations

Most of the work in cooperative mobile robotics has biological inspirations, imitating social characteristics of insects and animals after the introduction of the new robotics paradigm of behaviour-based control. This behaviour-based paradigm has had a strong influence in the design of the cooperative mobile robotics research. The most common application uses of the simple local control rules of various biological societies, particularly ants, bees, and birds, to the development of similar behaviours in cooperative robot systems. Work has demonstrated the ability for multi-robot teams to flock, disperse, aggregate, forage, and follow trails. The application of the dynamics of ecosystems has also been applied to the development of multi-robot teams that demonstrate emergent cooperation as a result of acting on selfish interests. To some extent, cooperation in higher animals, such as wolf packs, has generated advances in cooperative control. Significant study in predator-prey systems has occurred, although primarily in simulation. In [Vidal et al, 2002] which implements a pursuit-evasion task on a physical team of aerial and ground vehicles. They evaluate various pursuit policies relaying expected capture times to the speed and intelligence of the evaders and the sensing capabilities of the pursuers.

5.2 Communication, Architectures

The communication in multi-robot teams has been extensively studied. There are implicit and explicit communication, in which implicit communication occurs as a side-effect of other actions, whereas explicit communication is a specific act designed solely to convey information to other robots on the team. Several researchers have studied the effect of communication on the performance of multi-robot teams in a variety of tasks, and have concluded that communication provides certain benefit for particular types of tasks. Additionally, these researchers have found that, in many cases, communication of even a small amount of information can lead to great benefit. Other work in multi-robot communication has focused on representations of languages and the grounding of these representations in the physical world. This work has extended to achieving fault tolerance in multi-robot communication, such as setting up and maintaining distributed communications networks and ensuring reliability in multi-robot communications. In [Shen et al, 2002] communication enabling multi-robot teams to operate reliably in a faulty communication environment. In [Rybskiet al, 2002], explores communications in a teams of miniature robots that must use very low capacity RF communications due to their small size. They approach this issue through the use of process scheduling to share the available communications resources. [2] The distributed robotics has focused on the development of architectures, task planning capabilities, and control. This research area addresses the issues of action selection, delegation of authority and control, the communication structure, heterogeneity versus homogeneity of robots, achieving coherence amidst local actions, resolution of conflicts, and other related issues. The architecture, developed for multi-robot teams, tends to focus on providing a specific type of capability to the distributed robot team. Capabilities that have been of particular emphasis include task planning, fault tolerance, swarm control, human design of mission plans, role assignment, and so forth. [2]

5.3 Localization/Mapping/Exploration.

Research has been carried out in the area of localization, mapping, and exploration for multi-robot teams. Almost all of the work has been aimed at 2D environments. Initially, most of this research took an existing algorithm developed for single robot mapping, localization, or exploration, and extended it to multiple robots. Some algorithms have developed that are fundamentally distributed. One example of this work is given in [13], which takes advantage of multiple robots to improve positioning accuracy beyond what is possible with single

robots. Another example is a decentralized Kalman-filter based approach to enable a group of mobile robots to simultaneously localize by sensing their team-mates and combining positioning information from all the team members. They illustrate the effectiveness of their approach through application on a team of three physical robots. [Roumeliotis and Bekey , 2002].

Others works develops and analyzes a probabilistic, vision-based state estimation method that enables robot team members to estimate their joint positions in a known environment. Their approach also enables robot team members to track positions of autonomously moving objects. They illustrate their approach on physical robots in the multi-robot soccer domain.

Research approaches to localization, mapping, and exploration into the multi-robot version can be described using the familiar categories based on the use of landmarks, scan-matching, and/or graphs, andwhich use either range sensors (such as sonar or laser) or vision sensors. [2]

5.4 Object Transport And Manipulation

Enabling multiple robots to cooperatively carry, push, or manipulate common objects has been a long-standing, yet difficult, goal of multi-robot systems. Only a few projects have been demonstrated on physical robot systems. This research area has a number of practical applications that make it of particular interest for study. Numerous variations on this task area have been studied, including constrained and unconstrained motions, two-robot teams versus "swarm"-type teams, compliant versus non-compliant grasping mechanisms, cluttered versus uncluttered environments, global system models versus distributed models, and so forth. Perhaps the most demonstrated task involving cooperative transport is the pushing of objects by multi-robot teams. This task seems inherently easier than the carry task, in which multiple robots must grip common objects and navigate to a destination in a coordinated fashion. A novel form of multi-robot transportation that has been demonstrated is the use of ropes wrapped around objects to move them along desired trajectories. A research explores the cooperative transport task by multiple mobile robots in an unknown static environment. Their approach enables robot team members to displace objects that are interfering with the transport task, and to cooperatively push objects to a destination. In other presents a novel approach for cooperative manipulation that is based on formation control. Their approach enables robot teams to cooperatively manipulate obstacles by trapping them inside the multi-robot formation. They demonstrate their results on a team of three physical robots.

5.5 Motion Coordination

Another topic in multi-robot teams is that of motion coordination. Research themes in this domain that have been particularly well studied include multi-robot path planning, traffic control, formation generation, and formation keeping. Most of these issues are now fairly well understood, although demonstration of these techniques in physical multi-robot teams (rather than in simulation) has been limited. The motion coordination problem in the form of path planning for multiple robots is presented that performs path planning via checkpoint and dynamic priority assignment using statistical estimates of the environment's motion structure. Additionally, they explore the issue of vision-based surveillance to track multiple moving objects in a cluttered scene. The results of their approaches are illustrated using a variety of experiments. [2]

There exits different approaches for coordination of multiple robots, considering integration of communication constraints in the coordination of robots. In Yamauchi approach, uses a technique for multi-robot exploration which is decentralized, cooperative and robust to individual failures. He demonstrated a frontier-based exploration which can be used to explore office buildings. He used evidence grids which are Cartesian grids, which store the probability of occupancy of the space (prior probability equal 0, 5). The robots create their evidence grids by using their sensors, classifying each cell as Open, occupancy probability $<$ prior probability; Unknown, occupancy probability $=$ prior probability; and Occupied: occupancy probability $>$ prior probability. In this manner, any open cell which is near to an unknown cell is labelled as frontier edge cell. Frontier regions are formed by adjacent frontier edge cells.

The robots have to move to boundary between open space and unexplored part of the space to gain much new information. After a robot detected frontiers in the evidence grid, it tries to go to the nearest frontier. Besides, robots use path planner to find the nearest unvisited frontier and reactive obstacle avoidance behaviours to hinder collisions with unseen obstacles on the evidence grid.

After reaching and performing a 360 degree sensor sweep in the frontier, it adds the new information to the evidence grid of its local map. In multi-robot exploration, there is a local evidence grid which is available for all the robots. Besides that, every robot is creating its own global evidence grid. This global evidence grid shows its

knowledge about the environment for the robot. Using two separate evidence grid gives the advantage of being decentralized and cooperative. For instance, when a robot detects a new frontier, it starts to travel this point. After reaching this point, it performs a sensor sweep. By this way, it updates its local evidence grid with the new information. Moreover, it transmits updated local evidence grid information to the other robots. Besides that, global evidence grid is integrated with local evidence grid in a straightforward way. Using this cooperative approach, all new information is available for the other robots. Thus, each robot can update its own global evidence grid. There are two advantages of sharing a global map. Firstly, robots can make decision about which frontiers are unexplored yet by using updated maps. This improves the efficiency of exploration. Additionally, if a robot is disabled in the area, this won't affect the other robots. In his study, he developed a coordination approach based on frontiers. His frontier based approach is became a milestone for the following researches. [43]

The approach of R. Simmons, the coordination among robots is done to explore and create a map. This multi-robot exploration and mapping which is based on cost of exploration and estimation of expected information gain. They decreased the completion time of creating map task by keeping the robots well separated, resulting of the minimizing the overlap in information gain between the robots. Besides, they distributed most of the computation, which takes place in exploring and creating the map. Global map is constructed by the distinctive sensor information of the robots same as Yamauchi's approach. As a result, creating a consistent global map and assigning task to each robot which maximizes the overall utility are efficient examples of coordinated mapping and coordinated exploration, respectively. Besides, there are three important achievements with their approach: they used same software for both the local and global mapping; the robots update the global map with new information, even if they cannot communicate each other directly. This minimizes the alignment in local maps. Lastly, after each robot creates its local map, it sends local map to central mapper module. In the process of combining the maps to create one global map, central mapper module combines the data iteratively. Thus the localization error is minimizing. Their approach to distribute most of the computation among the robots is remarkable. However, they have two assumptions in this situation. Firstly, robots know their position relative to another. More sophisticated methods must be found for mapping and localization where the initial positions are not known by robots. Secondly, the researchers assumed that robots have an access to high-bandwidth communication, [36].

Another approach proposed, is based on separating the environments into stripes. These stripes show the successively explored environment by the multi-robots. However, in this approach, if one robot moves to a point, the rest of the robots wait on their position and watch for the moving robot. This approach significantly decreases odometry error, however it is not designed for distribution of the robots and the robots tend to stay close. [32]

The coordination of multi robots combines the global map; central mapper distributes the new map. Additionally, robots produce new bids from the updated map information. By using these bids, robots can mark the map cells as obstacle, clear or unknown. They used the frontier-based algorithm for exploration which is found by Yamauchi. However, there are two modifications to frontier-based approach. Firstly, they determined the estimation of the cost of travelling a frontier cell by calculating the optimal paths from the robots initial positions. These computations are made simultaneously by using a flood-fill algorithm. By determining the cost of travelling to a frontier cell for each robot, they assign this exploring task to the robot which has the optimal path. Secondly, they estimated information gain from the frontier cell by creating a rectangle which approximates the information gain region. Thus, executive uses these rectangles to estimate potential overlaps on coverage. By finding the cost of travelling and estimating the information gain in the frontier cells, executive assigns tasks to the robots. The idea behind the assigning tasks is discounting. For example, executive finds the bid location with the highest utility for a robot. After that, it discounts the bids of the remaining robots, selects another robot which has highest utility among the other robots. This task assignment continues till no robot or no task remains.

A new approach for coordination uses market-based approach, by minimizing the costs and maximizing the benefits. Like the previous approaches, robots communicate with each other continuously to receive new information about the environment. Thus, robots can improve their current plans. Even though there is a central agent, they are not dependent the central agent. However, in exploration process, if the central agent is reachable, the robots are communicating with central agent to learn if there are new goal points. [48]

Auction methods have been investigated as effective, decentralized methods for multi-robot coordination. Theoretical analysis and experimental of the performance of auction methods for multi-robot routing, has shown great potential. These methods are shown to offer theoretical guarantees for a variety of bidding rules and team objectives.

The problem of routing in multi-robot is specified by a set of robots, $R = \{r_1, r_2, \ldots, r_n\}$, a set of targets, $T = \{t_1, t_2, \ldots, t_m\}$, their locations, and a non-negative cost function $c(i, j), i, j \square R \square T$, which denotes the cost of moving between locations i and j. Assuming that these costs are symmetric, $c(i, j) = c(j, i)$, are the same for

all robots, and satisfy the triangle inequality. Travel distances and travel times between locations satisfy these assumptions in any typical environment. The objective of multi-robot routing is to find an allocation of targets to robots and a path for each robot that visits all targets allocated to it so that a team objective is optimized. In Auction methods the team objectives could be: MINISUM: Minimize the sum of the robot path costs over all robots, MINIMAX: Minimize the maximum robot path cost over all robots. MINIAVE: Minimize the average target path cost over all targets. The robot path cost of a robot r is the sum of the costs along its entire path, from its initial location to the last target on its path. The target path cost of a target t is the total cost of the path traversed by robot r from its initial location up to target t, where r is the unique robot visiting t. Optimizing performance for any of the three team objectives is NP-hard, considering that there is no polynomial time algorithm for solving multi-robot routing optimally with the MINISUM, the MINIMAX, or the MINIAVE objective, unless $P = NP$.

The main advantage of this multi-round auction mechanism is its simplicity and the fact that it allows for a decentralized implementation on real robots. Initially, each robot needs to know its own location, the location of all targets, and the number of robots (the number of bids in each round), but not the locations of the other robots. In each round, each robot computes its single bid locally and in parallel with the other robots, broadcasts the bid to the other robots, receives the bids of the other robots, and then locally determines the winning bid. This procedure is repeated in every round of the auction. Broadcasting can be achieved by means of relaying messages from robot to robot. Clearly, there is no need for a central auctioneer, and therefore, there is no single point of global failure in the system. Notice also the low communication complexity; each robot needs to receive n numbers (bids) in each of the m rounds, therefore O(nm) numbers need to be communicated over any single link. [18]

5.6 Reconfigurable systems

The motivation for reconfigurable distributed systems of this work is to achieve function from shape, allowing individual modules, or robots, to connect and re-connect in various ways to generate a desired shape to serve a needed function. These systems have the theoretical capability of showing great robustness, versatility, and even self-repair. Most of the work in this area involves identical modules with interconnection mechanisms that allow either manual or automatic reconfiguration. These systems have been demonstrated to form into various navigation configurations, including a rolling track motion, an earthworm or snake motion, and a spider or hexapod motion. Some systems employ a cube-type arrangement, with modules able to connect in various ways to form matrices or lattices for specific functions. An important example of this research is in [Shen et al, 2002] that presents a biologically inspired approach for adaptive communication in self-reconfigurable and dynamic networks, as well as physical module reconfiguration for accomplishing global effects such as locomotion.

6 Colony of Ant robots.

Social insects that live in colonies, such as ants, termites, wasps, and bees, develop specific tasks according to their role in the colony. One of the main tasks is the search for food. Real ants search food without visual feedback (they are practically blind), and they can adapt to changes in the environment, optimizing the path between the nest and the food source. This fact is the result of stigmergy, which involves positive feedback, given by the continuous deposit of a chemical substance, known as pheromone. Ants are social insects capable of short-range interactions, yet communities of ants are able to solve complex problems efficiently and reliably. Ants have, therefore, become a source of algorithmic ideas for distributed systems where a robot (or a computer) is the "individual" and a swarm of robots (or the network) plays the role of the "colony."

Ant robots are simple and cheap robots with limited sensing and computational capabilities. This makes it feasible to deploy teams of ant robots and take advantage of the resulting fault tolerance and parallelism. They cannot use conventional planning networks due to their limitation and their behaviour is driven by local interactions. Ant Robots almost never know exactly where they are in the environment.

A common way is to use probabilistic planning, provides to robots, the best possible location estimate, to achieve their goals without ever worrying about where they are in the terrain. Other approach of Ant robots can communicate via markings that they leave in the terrain, similar to ants that lay and follow pheromone trails, and solving robot-navigation tasks in a robust way. Using Pheromone Traces of alcohol [Sharpe et al.], heat [Russell], odor [34], and virtual traces [Vaughan et al.; Payton et al.], no location is estimates, no planning is need, no direct communication with a simpler hardware and software. The result is a very robust navigation. It has been developed a theoretical foundation for ant robotics, based on ideas from real-time heuristic search, stochastic analysis, and graph theory. [39], [17]

Teams of robots can do mine sweeping, surveillance, search-and-rescue, guarding, surface inspection and many others work. For example, a team of robots that cover terrain repeatedly can guard a museum at night. [35] The main areas, involved to have group of robots are: Agent coordination (swarms), Robotics (robot architectures, ant robots, sensor networks), Search (real-time search), complexity analysis of graph algorithms, Communication.

6.1 Ant colony optimization

The ants construct a pheromone trail in the search for a shorter path from nest to the food. When an obstacle is inserted in the path, ants spread to both sides of the obstacle, since there is no clear trail to follow. As the ants go around the obstacle and find the previous pheromone trail again, a new pheromone trail will be formed around the obstacle [Colorni et al., 1991]. This trail will be stronger in the shortest path than in the longest path. As shown in [Parpinelli et al., 2002], there are many differences between real ants and artificial ants, mainly: artificial ants have memory, they are completely blind and time is discrete. On the other hand, an ant colony system allows simulation of the behaviour of real-world ant colonies, such as: artificial ants have preference for trails with larger amounts of pheromone, shorter paths have a stronger increment in pheromone, and there is an indirect communication system between ants, the pheromone trail, to find the best path.

The term collective behaviour, in the robotics literature, means: joint collaborative behaviour that is directed toward some goal in which there is a common interest; a form of interaction, usually based on communication; and joining together for doing something that creates a progressive result such as increasing performance or saving time. Cooperative behaviour is to associate with another or others for mutual, often economic, benefit.

A collaborative robot is a robot designed, for example, to assist human beings as a guide or assistor in a specific task. A cooperative robots are a group of robots that can work together to move large objects, sharing the load. Having coordination in actions, sharing sensors and computing power, multi robots can perform tasks such as drill holes and pitch tents in tight coordination. They can carry out the tasks in an unstructured outdoor environment.

7 Mutiagents Systems applied to Colony of Robots.

Using decentralized approaches, as Multi-Agent Systems, gives a new way for modelling a colony of robots. This approach is able to generate a self-organized system with some robust and efficient ways of solving problematic like some insect societies, as ants. [34]

This focus provides three multi-level advances:

(i) the development of the concept of Complex Systems give a new way of modelling, based on decentralized representation composed of interaction network of entities from where emergent properties appear ;

(ii) Object-Oriented programming had proposed a first step in the decomposition of computing and so in the following, agent oriented programming adds to objects some autonomous properties;

(iii) the development of huge computer networks promote the distributed computing which finally allowed implementations of these previous concepts.

[46]. Complex systems are usually presented as some systems of interacting entities which can be represented as a kind of networks. Agent-based can represent the interactions between entities as communication processes.

7.1 Robot Architecture.

Robot architecture has been proposed, based on a multi-agent system (MAS). The agents have the goal: to control the robot and to do it intelligent, while competing for resources. This approach produce a more robust, flexible, reusable, generic and reliable architecture that can be easily modified and completed to permit social behaviour among robots; it is also holonic multi-agent systems. The agents that make up the proposed architecture may also be Multiagent systems themselves. The Task Planning Agent is a multi-agent system formed by planner agents and a coordination agent. [21]

7.2 Multiagent plan coordination.

The Multiagent plan coordination problem arises whenever multiple agents plan to achieve their individual goals independently, but might mutually benefit by coordinating their plans to avoid working at cross purposes or duplicating effort. Although variations of this problem have been studied, there is no agreement over a general characterization of the problem. A general framework that extends the partial order, causal-link plan representation to the Multiagent case, and that treats coordination as a form of iterative repair of plan flaws that cross agents. Multiagent planning has acquired a variety of meanings over the years. In part, this may be due to the ambiguity of exactly what someone considers to be "multiagent" about the planning. In some work, it is the planning process that is multiagent; for example, multiple agents, each with specialized expertise in certain aspects of planning, might collaborate to formulate a complex plan that none of them could have generated alone. In other work, it is the product of planning, the plan itself, that is multiagent, in the sense that it specifies the activities of multiple actors in the environment such that they collectively achieve their individual and/or common goals. And sometimes, it is both where multiple agents interact to arrive at plans that will be carried out by multiple (and often, it is implicitly assumed, the same) agents. In the third class of problems, a Multiagent Plan Coordination Problems (MPCPs), in which multiple agents, has each plan the individual activities, but might mutually benefit by coordinating their plans to avoid interfering with each other unnecessarily duplicating effort. Multiagent plan coordination differs from "team planning," in which agents must work together more tightly as a team in order to achieve their joint goals. Instead, multiagent plan coordination is suited to agents that are loosely-coupled (nearly independent), where each agent can achieve its own goals by itself, but the presence of other agents who are also asynchronously operating in the same environment leads to potential conflicts and cooperative opportunities. [8] Other similar approaches is described, where the problem of finding plans with minimal make span is considered. In both, the degree of coupling measures, the degree of interaction between different plans (threads) and thus affects the inherent difficulty of the planning problem. In general, the multiagent plan coordination problem is known to be NP-Hard. It has developed a rigorous computational theory of single-agent and multi-agent plan coordination, and implemented an efficient and optimal algorithm that, under assumed characteristics, is polynomial with respect to the size of the plan coordination problem. [44, 30]

The autonomy of colony of robots depends on the behaviour of each agent associated to each robot in terms of actions and in terms of flexible group decision making. To achieve this objective, it is necessary that agent architectures can help in designing the architecture of the software of each robots; a multi-agent approach can address the problem of interactions between robots; and automated planning can provide the basis of robots intelligence.

A vehicle can be adapted to act as robots. It has proposed architecture and distributed planning method for multi-vehicle missions contribute to the increase of vehicle intelligence and autonomy. The integration of online planning, disruptive events in absence of human intervention do not lead necessarily to aborting the mission. However, it is important to note that the architecture addresses a specific class of multi-vehicle missions. For this class the plan exists at the beginning of the mission and provides actions up to the end of the mission. In a context where there is a large uncertainty about the ending conditions of the mission or where there are systematically a large difference between the situation expected at planning time and the actual situation, other architectures based on a more systematic activation of the planning module are more suited.

7.3 Cooperative Planning.

The problem of the cooperative planning is very complex. To specify a planning task for the system of entities more precisely, it has to know about how the system shares the knowledge; how precise is the map of the working environment; when the environment is static or changing; and the kind of a task is being solved. When system has multiple entities, it can be organized in a centralized way, or a decentralized. Centralized systems have a central control unit managing tasks for other entities, the knowledge about the solved tasks and working environment, system configuration, actual state of the system (positions of particular entities) is stored and maintained in the central unit. It also distributes local tasks according to priorities to other entities. On the other hand, knowledge in decentralized systems is shared among all entities where each entity plans and performs its own activities autonomously with respect to needs and request of other entities. Advantages of centralized systems are in terms of traffic control, resource management, and task optimization. On the other hand decentralized systems are superior in terms of robustness and scalability of the systems. Robustness can be defined as the ability of a system to gracefully degrade when some entity in the system fails. Robust systems are able to work properly even in the case that any entity is malfunctioning, as long as there are some functioning robots. Centralized systems do not have this characteristic because failure of the central unit disables the entire system. It can take advantage of both systems,

using hybrid solutions. [47]

7.4 Planning task

Planning task is the amount of information about working environment. If the map of working area is available and accurate, the planning problem leads to geometrical optimization, computed off-line. When robots are operating in unknown environments, the task cannot be divided optimally without complete information. This problem is solved on-line because new information about environment can be obtained during fulfilling the plan. Another situation sets in when a map is available but the working environment is rapidly changing or a map is not precise. In this case, primary activities can be planned off-line based on the available map. Concrete actions can be specified on-line in more details with respect to acquired information about the environment in which the system operates.

If the colony of robots has only three kinds of tasks such as coverage problem: exploration, and coordinated planning. Coverage planning for a team of robots deals with the problem ensuring that every point in the working environment is visited by at least one robot. Coverage planning can be used for a number of different tasks, for example floor cleaning, grass cutting, foraging and mine detection and removal. Exploration of a working area is a similar task; the goal is not to "reach" all places in the environment, but to "see" all places. The main aim during exploration is how to move particular robots in order to minimize the time needed to completely explore the environment. When mobile robots have the ability to explore a surrounding environment efficiently, they are able to build a model (map) of their environment and to solve more complex tasks that ensue from mapping like the detection of specified objects. The key question of coordinated planning is how to plan paths for particular robots in order to avoid collisions. In other words, two robots cannot be at the same place in the same time.

7.5 Planning multiple robots.

Planning for multiple robots is one of the main research topics within multi-robots systems [28]. Differences in the approaches are mainly in methods of knowledge sharing. Multi agent approaches and techniques are applicable (mainly in task of the distribution for multiple entities) but it is necessary to consider a low preciseness of localization and mapping strategies. An example of the application of the behaviour-based multi agent approaches with consideration to physical multi-robot systems is presented in [12].

The common attribute of all planning problems is the effort to find the optimal solution. A typical criterion to be optimized is the overall time spent by all team members during the task execution or the sum of lengths of particular robot's paths. The type of criterion leads to applicability of different strategy planning for multiple robots. Basic problem of the planning is the path planning. The research for robot's path planning has centred on the problem of finding a path from a start location to a goal location, while minimizing one or more parameters such as length of path, energy consumption or journey time. The optimal path planning is the essential problem of the exploration and the coverage task. [29]

7.6 Multiagent Learning.

Multi-Robot Systems MRS can often be used to fulfil the tasks with uncertainties, incomplete information, distributed control, and asynchronous computation, etc. The performance of MRS in redundancy and co-operation contributes to task solutions with a more reliable, faster, or cheaper way. Multiagent reinforcement learning can be useful for multi-robot systems The challenges in MRSs involve basic behaviours, such as trajectory tracking, formation-keeping control, and collision avoidance, or allocating tasks, communication, coordinating actions, team reasoning, etc. For a practical multi-robot system, firstly basic behaviours or lower functions must be feasible or available. In upper modules, for task allocation and planning, have to be designed carefully. Robots in MRSs have to learn from, and adapt to their operating environment and their counterparts. Thus control and learning become two important and challenging problems in MRSs. [24]

Multiagent reinforcement learning RL allows participating robots to learn mapping from their states to their actions by rewards or payoffs obtained through interacting with their environment. Robots in MRSs are expected to coordinate their behaviours to achieve their goals. These robots can either obtain cooperative behaviours or accelerate their learning speed through learning. [45]

Among RL algorithms, Q-learning has attracted a great deal of attention. Explicit presentation of an emergent idea of cooperative behaviours through an individual Q-learning algorithm can be found. Improving learning efficiency through co-learning was shown. The study indicates that K co-operative robots learned faster than

they did individually. It has also demonstrated that sharing perception and learning experience can accelerate the learning processes within robot group.

Recently there has been growing interests in scaling multiagent RL to MRSs. Although RL seems to be a good option for learning in multiagent systems, the continuous state and action spaces often hamper its applicability in MRSs. Fuzzy logic methodology seems to be a candidate for dealing with the approximation and generalization issues in the RL of multiagent systems. However, this scaling approach still remains open. [20]

7.7 Ontology and Semantic.

Several methodologies exist for building multiagent systems; few of them address the information domain of the system. Just as important as the system's representation of the information domain is the various agents' information domain view. Heterogeneous systems can contain agents with differing data models, a case that can occur when reusing previously built agents or integrating legacy components into the system. Most existing methodologies lack specific guidance on the development of the information domain specification for a multiagent system and for the agents in the system.

An appropriate methodology for developing ontology's must be defined for designers to use for specifying domain representations in multiagent systems. The existing methodologies for designing domain ontologies are built to describe everything about a specific domain; however, this is not appropriate for multiagent systems because the system ontology should only specify the information required for proper system execution. The system ontology acts as a prerequisite for future reuse of the system, as the ontology specifies the view of the information domain used by the multiagent system. Any system that reuses the developed multiagent system must ensure that the previously developed system ontology does not conflict with the ontology being used in the new system.

Once the system ontology is constructed, a multiagent system design methodology should allow the analyst to specify objects from the data model as parameters in the conversations between the agents. To ensure the proper functionality of the multiagent system, the designer must be able to verify that the agents have the necessary information required for system execution. Since the information is represented in the classes of the data model, the design of the methodology must show the classes passed between agents. [22]

7.8 Multiagent Systems Engineering (MaSE)

Multiagent Systems Engineering is an attempt to answer the sixth challenge, how to engineer practical multiagent systems, and to provide a framework for solving the first five challenges. It uses multiagent systems for developing intelligent, distributed software systems. MaSE uses two languages to describe agents and multiagent systems: the Agent Modelling Language (AgML) and the Agent Definition Language (AgDL), to define a methodology specifically for formal agent system synthesis. Both AgML and AgDL will be defined with a precise, formal semantics. The methodology can also be successfully applied with traditional software implementation techniques as well.

There are a lists six challenges of multiagent systems: to decompose problems and allocate tasks to individual agents. To coordinate agent control and communications. To make multiple agents act in a coherent manner. To make individual agents reason about other agents and the state of coordination. To reconcile conflicting goals between coordinating agents. How to engineer practical multiagent systems. AgML and AgDL semantics are based of multi-sorted algebras. Algebraic approaches have the advantage that there has been a great deal of work in automatically synthesizing code from algebraic specifications. The work is similar in many respects to the agent methodologies based on object-oriented concepts.

However, few of these have a formal basis. Some work in formalization of agent systems has been performed in but has focused on formal modelling and not automated code synthesis. MaSE and AgML together provide many advantages over traditional software engineering techniques. Because of this abstraction, MaSE can capture traditional object-oriented systems as well as agent-based systems for which traditional techniques are inappropriate. Then, it has a more concise representation than object-oriented techniques. It has a formal syntax and semantics. [38] [20]

8 Summarizing Colony of Robots.

In this paragraph is presented a synthesis what would be a Colony of Robots, its characteristics, its components, its applications, its basis and the way to build the Colony as an Engineering Project.

8.1 Parts of a Colony of robots.

A colony could have the following components and actors: Centre of Colonies, Nest of the Colony, Colony Leader, Agency Leader, and different types of Working Robots.

Centre of Colonies is a place controlled by human beings. It is located far from zone of work of the colony. Persons determine the works that the Colony must realize, when initial and final time, in what place, which it is necessary to do and when the Colony must report its work. The orders of work are sent to the nest of the Colony in a remote way.

Nest of the Colony is where the Colony is placed, composed of several heterogeneous autonomous robots, which has a Colony Leader and several Agency Leaders who assign works.

Colony Leader is a robot that has communication with the Centre of Colonies. This Leader receives the orders of work from Centre of Colonies; it sends reports of realized work, problems happened in the work or in the colony; it decomposes orders into tasks and assigns a task to the different Agency Leaders Leading. It receives the state of works and report s to Agency Leaders.

Agency Leader, are robots that have communication with the Colony Leader and with Working Robots or Robot Agents. It receives orders, from Colony Leader, of the task that it is necessary to realize, determines that the needs of Working Robots for the task and sends orders to go to the place of work and to do the task.

Working Robots are a group of heterogeneous mobile robots that receives orders Agency Leader, to move to a workplace to carry out a given task. The Working Robots communicate themselves to determine the best path to arrive to the workplace, to warn if they need help to do a task, to report the work realized, to report when it has faults or it has little energy.

The Working Robots also communicate with an Agency Leader, reporting the works done, the problems of the group of robots and the need of help, etc. It could be some types of these robots are specialist. One of them can have computer vision systems; others it have manipulation systems of objects; another ones with derricks systems to transport robot; robots repairers, etc.

8.2 The Characteristics of the Colony.

A Colony of Robots would have to have defined its characteristics. The main characteristics can selected for the persons in the Centre of Colonies. Normally, the Colony characteristics must be: Model of the Colony; Operational Environment; Communication in the Colony; Colony Coordination; Cooperative and Collaborative Work, among the Working Robots to carry out works; Robustness, Fault tolerance, Reliability, Flexibility, Adaptively and Coherence; Reconfigurability, Localization, Mapping, Exploration, Object transport and manipulation.

Model of the Colony. The colony must have a biological inspiration model of how operating. It has to have a model to imitate, such as animal colony or of insects colony. There are several models developed, usually it is an Ant Colony Model.

Operational Environment. The colony of robots must be designed for thinking in the environment that it will settle. According to this way, different types of robots are needed. If an average normal environments, mobile robots need to be designed to realize works in that environment. If they are extreme environments, the robots must be designed for those conditions. An alternative is “to automate” a vehicle that already exists, in the way of which it could be autonomous and apt to do the tasks that it are wanted.

Communication in the Colony. The communication is important in a Colony. These communications have to be reliable, precise and fault tolerant. The communications are among the Centre of Colonies and the Colony; among Colony Leader and Agency Leaders; among these Agency Leaders and the Working Robots; and among Working Robots.

Colony Coordination. Coordination must exist between the different Working Robots of the Colony for the movement to the work place; in the cooperation and collaboration among robots. In general, it is desired that the colony be capable of coordinating path planning, traffic control, formation generation, and formation keeping.

Cooperative and Collaborative Work, among the Working Robots to carry out works. Robustness, Fault tolerance, Reliability, Flexibility, Adaptively and Coherence in all the different robots in the Colony.

Reconfigurability. The colony must be reconfigurable distributed systems to achieve function from shape, allowing robots to connect and re-connect in various ways to generate a desired shape to serve a needed function. These systems have to have capability of showing great robustness, versatility, and even self-repair.

Localization, Mapping, Exploration. The colony of robots must be capable of knowing the coordinates where it is and where that it must go, using GPS systems (Global Positioning System). It has to be capable of creating a

map of the tour and of being able to explore. This information must be shared for groups of robots involved in a work.

Object transport and manipulation the colony must be capable of manipulating, transporting or pushing simple objects, in individual way or in collaboration and / or cooperation way, with other robots.

8.3 Colony wanted and for what.

What is expected from a colony is that it realizes the works assigned in an efficient, reliable form and it reports the works done, achievements and problems. The works can be very diverse depending what the persons want in the Centre of Colonies. The Colony can be utilized in Exploration, in Industries, at Home, in Military Forces, in Education, and many others uses.

In Exploration. When a Colony of Robots has assigned works to know certain area, the exploration has to know, for example, temperature, pressure, ozone level, radioactivity, etc. In general, it is needed to measure physical or chemical variables, to inspect the place, searching path, to make images of the environment, animals, vegetables, to see the composition of the area, water, etc. This exploration can realize in different locations to measure the environment, especially if it is hostile (volcanoes, deserts, Antarctic, Arctic, etc.). Other kind of exploration is the spatial exploration to know the Moon and Mars.

In Industries. There can have groups of robots to do maintenance of pipelines, for painting, for making measurements and to do works in dangerous zones, etc. Also, the group can be useful in flexible Manufacturing Systems FMS; in the agricultural industry to detect plagues of insects, to determine when it is necessary to harvest, etc.; in industries of production of raw materials and production of energy.

At homes. To improve quality of live of the persons. For example groups of robots to sweep, to clean, to paint etc.

In Education; in Military Forces and many Others.

8.4 How the Colony of Robots is built.

The colony of robots is constructed from the specifications given by the users. Though the principles are the same, the design of the Centre of the Colony, the Nest of the Colony, the different robots, they depend on the works to be realized, on the environment in which they are going to work and the desired precision. According these parameters and others specifications, it is necessary to do an Engineering Project to have the Colony of Robots needed. This Project includes the support of providing companies of Robotics, Automation, Communications, Electronic, etc.

8.5 Colony of Robots Background.

The Colonies of Robots base on physical principles, in general on scientific principles, and researches on Robotics, Computer Science, Mechanical, Automatic, Communications, Optimization and Planning, and many other fields related, shown in this paper.

8.6 Research in Colonies of Robots.

Though, there exist several groups doing research on colonies of robots, the motivations can be very different. Some Universities and Research Centres do works applicable to a Colony of Robots. It be better to have a multidisciplinary teams to creating a Colony of Robots.

9 Summary and Conclusions

In this paper has been presented an overview in colony of robots, considering models of communities of robots, the behaviour of groups of robots, ant robots colony, communication inside de colony, multirobots characteristics and multiagents systems applied to multirobots.

To have control of a colony of robots that working together in a collaborative and cooperative way in a non structured environments, is important considerer the communication among the robots, the way of planning and coordination of the robots, how the object has transport and manipulation, and reconfiguration. The colony has to have a cooperative architecture with robustness, a good fault tolerance, to be reliability, to present flexibility,

adaptively and coherence. Communication among robots of the colony permits to have a distributed control, planning and coordination to perform the tasks assign for achieving the aims of the colony.

It is expected that the colony of robots has a behaviour as insects, with intelligent distributed behaviours that can do well the tasks, in a cooperative and collaborative way. This colony could have low-cost robots, heterogeneous architecture and distributed algorithms. The kinds of tasks of the colony of robots normally are exploration, and coordinated planning. To exhibit intelligence, the robot architecture in the colony, is based on a multi-agent system. In this manner, is possible to have control of each robot in an intelligent, way.

This approach produce a more robust, flexible, reusable, generic and reliable architecture that can be easily modified and completed to permit social behaviour among robots. Each robot is a Multiagent system to be more efficient. Multiagent planning and multiagent coordination inside the colony are good solutions. The applications of colonies of robots several fields in industries, scientific exploration and at home. A guideline what considering in create a Colony of Robots has been presented.

References

[1] Ali Umut Irturk, Distributed Multi-robot Coordination For Area Exploration and Mapping. University of California Santa Barbara, 2006

[2] Tamio Arai, Enrico Pagello, Lynne E. Parker. Advances in Multi-Robot Systems. IEEE Transactions on Robotics and Automation, vol. 18, no. 5, pp. 655-661. October 2002.

[3] H. Asama, K. Ozaki, A. Matsumoto, Y. Ishida, and I. Endo. Development of task assignment system using communication for multiple autonomous robots. Journal of Robotics and Mechatronics. 4(2):12-127, 1992.

[4] Alan Bond and Less Gasser, Readings in Distributed Artificial Intelligence. Morgan Kaufmann, 1988.

[5] Rodney A. Brooks, Pattie Maes, Maja Mataric, and Grinnell Moore. Lunar base construction robots. In Proceedings of the IEEE International Workshop on Intelligent Robots and Systems, pages 389-392, Tsuchiura, Japan, 1990.

[6] Philippe Caloud, Wonyun Choi, Jean-Claude Latombe, Claude Le Pape, and Mark Yim. Indoor automation with many mobile robots. In Proceedings of the IEEE International Workshop on Intelligent Robots and Systems, pages 67-72, Tsuchiura_ Japan, 1990.

[7] Paul Cohen, Michael Greenberg, David Hart, and Adele Howe. Real-time problem solving in the Phoenix environment. COINS Tech.l Report, University of Massachusetts at Amherst, 1990.

[8] Jeffrey S. Cox and Edmund H. Durfee, An Efficient Algorithm for Multiagent Plan Coordination. AAMAS'05, July 2529, 2005, Utrecht, Netherlands.

[9] Cortes, J. Martínez, S. Karatas, T. and Bullo, F. 2002. Coverage Control for Mobile Sensing Networks. In Proceedings of the IEEE Conference on Robotics and Automation, 1327-1332. Arlington, VA.

[10] Scott A. DeLoach, Multiagent Systems Engineering: A Methodology And Language for Designing Agent Systems. Agent-Oriented Information Systems (AOIS) '99

[11] Felix Duvallet, James Kong, Eugene Marinelli, Kevin Woo, Austin Buchan, Brian Coltin, Christopher Mar, Bradford Neuman. Developing a Low-Cost Robot Colony, AAAI Fall Symposium 2007 on Distributed Intelligent Systems

[12] M. S. Fontan and M. J. Mataric. Territorial multi-robot task division. IEEE Transactions of Robotics and Automation, 14(5), 1998.

[13] Fox, D. Burgard, W. Kruppa, H. and Thrun, S. 2000. A probabilistic approach to collaborative multi-robot localization. Autonomous Robots 8(3).

[14] Howard, A. Parker, L. and Sukhatme, G. 2006. Experiments with a Large Heterogeneous Mobile Robot Team: Exploration, Mapping, Deployment and Detection. The International Journal of Robotics Research 25: 431-447. 2006.

[15] Heger, F., and Singh, S. 2006. Sliding Autonomy for Complex Coordinated Multi-Robot Tasks: Analysis and Experiments. In Proceedings, Robotics: Systems and Science, Philadelphia.

[16] Hundwork, M. Goradia, A. Ning, X. Haffner, C. Klochko, C. and Mutka, M. 2006. Pervasive surveillance using a cooperative mobile sensor network. In Proceedings of the IEEE International Conference on Robotics and Automation.

[17] S. Koenig, B. Szymanski and Y. Liu. Efficient and Inefficient Ant Coverage Methods. Annals of Mathematics and Artificial Intelligence, 31, 41-76, 2001.

[18] Michail G. Lagoudakis , Evangelos Markakis, David Kempe, Pinar Keskinocak , Anton Kleywegt, Sven Koenig, Craig Tovey , Adam Meyerson , and Sonal Jain. Auction-Based Multi-Robot Routing. 2006.

[19] Y. Liu and S. Koenig. An Exact Algorithm for Solving MDPs under Risk-Sensitve Planning Objectives with One-Switch Utility Functions. In Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2008.

[20] Ellips Masehian, and Davoud Sedighizade, Classic and Heuristic Approaches in Robot Motion Planning - A Chronological Review. Proceedings of World Academy of Science, Engineering And Technology Volume 23 August 2007

[21] McFarlane, D.C. Bussmann, S.: Developments in Holonic Production Planning and Control. Int. Journal of Production Planning and Control,vol 11, N 6, pp 5522-536, 2000.

[22] D.L., Fikes, R. Rice, J. and Wilder, S. (2000). An Environment for Merging and Testing Large Ontologies. Principles of Knowledge Representation and Reasoning: proceedings of the Seventh International Conference. A. G. Cohn, F. Giunchiglia and B. Selman, editors. San Francisco, CA, Morgan Kaufmann Publishers. 2000

[23] McGuinness, D.L. and Wright, J. Conceptual Modeling for Configuration: A Description Logic-based Approach. Artificial Intelligence for Engineering Design, Analysis,and Manufacturing - special issue on Configuration. 1998

[24] J. Melvin, P. Keskinocak, S. Koenig, C. Tovey and B. Yuksel Ozkaya. Multi-Robot Routing with Rewards and Disjoint Time Windows. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS), 2332-2337, 2007.

[25] McLurkin, J. 2004. Stupid Robot Tricks: A Behavior-Based Distributed Algorithm Library for Programming Swarms of Robots. M.S. diss., Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Mass.

[26] Fabrice R. Noreils. Toward a robot architecture integrating cooperation between mobile robots. Application to indoor environment. The International Journal of Robotics Research, 12(1), 79-98, February 1993.

[27] Lynne E. Parker, Heterogeneous Multi-Robot Cooperation. MIT 1994.

[28] L. E. Parker, "Current State of the Art in Distributed Autonomous Mobile Robotics", in Distributed Autonomous Robotic Systems., L. E. Parker, G. Bekey, and J. Barhen eds., Springer-Verlag Tokyo 2000, pp. 3-12.

[29] Building Presence through Localization for Hybrid Telematic Systems, Research and task analysis of telematic planning. Report of Project funded by the European Community under the IST programme: Future and Emerging Technologies, PELOTE, IST-2001-38873, 2003.

[30] F. Pecora, R. Rasconi, and A. Cesta Assessing the bias of classical planning strategies on makespan-optimizing scheduling. In Proceedings of the 16th European Conference on Artificial Intelligence (2004), pp. 677-681.

[31] A. Portmann. Animals as Social Beings. The Viking Press, New York. 1961.

[32] I. Rekleitis, G. Dudek, and E. Milios, Multi-robot exploration of an unknown environment, efficiently reducing the odometry error. IJCAI Intert. Conference in AI, vol. 2, 1997.

[33] Rothenfluh, T.R. Gennari, J.H. Eriksson, H. Puerta, A.R. Tu, S.W. and Musen, M.A. (1996). Reusable ontologies, knowledge-acquisition tools, and performance systems: PROTÉGÉ-II solutions to Sisyphus-2. International Journal of Human-Computer Studies 44: 303-332.

[34] S. Russel and P. Norvig, "Artificial Intelligence, a modern approach", Prentice Hall, 2nd ed., 2003

[35] R. Simmons and S. Koenig. Probabilistic Robot Navigation in Partially Observable Environments. In Proceedings of the International Joint Conference on Artificial Intelligence, 1080-1087, 1995.

[36] R. Simmons, D. Apfelbaum, W. Burgard, D. Fox, M. Moors and S. Thrun, and H., Y. (2000), "Coordination for multi-robot exploration and mapping", In Proc. f the National Conference on Artificial Intelligence (AAAI).

[37] Sugar, T. G., and Kumar, V. C. 2002. Control of Cooperating Mobile Manipulators. In IEEE Transactions on Robotics and Automation, Vol.18, No.1, 94-103.

[38] Sycara, K. P. 1998, Multiagent Systems. AI Magazine 19(2): 79-92.

[39] J. Svennebring and S. Koenig. Trail-Laying Robots for Robust Terrain Coverage. In Proceedings of the International Conference on Robotics and Automation, 2003.

[40] Trebi-Ollennu, A. Nayar, H.D. Aghazarian, H. Ganino, A. Pirjanian, P. Kennedy, B. Huntsberger, T. and Schenker, P. 2002. Mars Rover Pair Cooperatively Transporting a Long Payload. In Proceedings of the IEEE International Conference on Robotics and Automation.

[41] N. Tinbergen. Social Behavior in Animal. Chapman and Hall Ltd. Great Britain, 1965.

[42] E. Wilson. The Insect Societies. The Belknap Press. Cambridge, 1971.

[43] B. Yamauchi, "Frontier-based exploration using multiple robots," in Proc. of the second International Conference on Autonomous Agents, Minneapolis, MN, USA, 1998, pp. 47-53

[44] Yang, Q. Intelligent Planning. Springer-Verlag, Berlin, 1997.

[45] Erfu Yang and Dongbing Gu Multiagent Reinforcement Learning for Multi-Robot Systems: A Survey. Engineering and Physical. 2006.

[46] M. Wooldridge, "An introduction to MultiAgent Systems", John Wiley and Sons, LTD, 2002

[47] X. Zheng and S. Koenig. Reaction Functions for Task Allocation to Cooperative Agents. In Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2008.

[48] R.M. Zlot, A. Stentz, M.B. Dias, and S. Thayer, "Multi-Robot Exploration Controlled By A Market Economy", IEEE International Conference on Robotics and Automation, May, 2002.

[49] M. Wooldridge and N. Jennings, 1995. IntelligentAgents: Theory and Practice. Knowledge Engineering Review, 10(2): 115-152.

Gastón Lefranc
Pontificia Universidad Católica de Valparaíso
Escuela de Ingeniería Eléctrica
Avenida Brasil 2147, Valparaíso, Chile
E-mail: glefranc@ieee.org

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 108-115

A Model of the Student Behaviour in a Virtual Educational Environment

Parallel session invited paper

Ioana Moisil

Abstract: In this paper I am presenting a behavioural model of the student as the main actor of a virtual learning environment. The model is part of a larger project - DANTE - *Socio-Cultural Models implemented through multi-agent architecture for e-learning* - that has as main objective the development of a model for the virtual education system, student centred, that facilitates the learning through collaboration as a form of social interaction. The project presumes the combination of the artificial intelligence (multi agent) system with elements of the socio-cultural theory of learning by collaboration and human actors. The model requires its own universe in which the human agents interact with the artificial ones (software agents). Student behaviour is modelled as a function of beliefs, affects, learning styles, intention and motivation, taking into account the context influences. The model is tributary to traditional behaviour and learning styles models (Rogers, Jung, Piaget, Fishbein, Kolb and followers). The model has been validated on a sample of 450 students at the university (18-22 years old, computer science and informatics specialities).
Keywords: e-learning, behaviour theories, learning style, socio-cultural theory, student model, student attitude, student motivation, Fishbein's reason action model, belief-importance model.

1 Background

The new information technologies offer to education many opportunities, but much more challenges. In order to obtain the maximum benefits and to answer to the main challenges educational software designers have to approach several learning models, to apply design methodologies centered on the student, to adapted to different learning styles, to different knowledge backgrounds. Many studies have shown that it is a fruitful idea (more flexibility, greater usability) to model human and artificial agents using similar learning paradigms, avoiding dangerous equivalence. Also using interaction design is important, when not only training but learning is the target, for it allows keeping alive the curiosity of the students, to motivate them. Being deeply involved in the design of software that they will use, students are also developing a certain sense of ownership that gives value to the content. For educational software designers but in the same degree for educators, the most important challenge is generated by e-learning. E-Learning- electronic learning - has appeared on the educational stage as a true successor of the 20th century paradigms of *Computer Aided Learning (CAL), Computer Aided Training (CAT) and Computer Aided Instruction*(CAI), encompassing *Computer Based Learning (CBL), Computer Based Training*(CBT) and all forms of web based learning. It was, and it is, a natural evolution product generated by technological advances. We are witnessing today a proliferation of e-learning products, most addressed to continuing education, distance learning, but also for students in schools and universities. Generally speaking, we can consider e-learning as a term used to refer computer enhanced learning [...], but when coming to define it, we must take care to specify the context in which it is used. Moreover, as technology is so rapidly evolving, e-learning is now accompanied by *M-learning* that is e-learning using mobile devices and technologies. What must be recognised is that e-learning is mainly about *learning* and the "e" be it for *electronic or enhancing or even empowered or enabled*, it is not the target. E-Learning has to be centred on people that are learning. It has to address different learning styles, different levels of basic knowledge. The technical aspects of electronic delivery of knowledge are, of course, very important, but they must be adapted to the learner and her/his environment. It is obvious that students learn better if the text books are well organized, richly illustrated, with clear headings, etc. And if we can add to all these animation and colours, the results are encouraging. E-Learning applications are manipulating a lot of different learning objects, from courses to projects and home works. The efficiency and the efficacy of these systems depend greatly on how well they adapt to the individual student profile. Many critics have attacked these products for their low psycho-pedagogical validity and a lack of standard quality assessing criteria. Challenges that e-learning software developers have to cope with are linked to these psycho-pedagogical and social characteristics that depend on the student's individuality. Learning Management Systems (LMS) offer facilities for managing authors, tutors, administrators by maintaining password systems and catalogues with roles, functions for controlling access to content, but they have a very few

options for monitoring students evolution (in general only quizzes and multiple choice grids for evaluation) and deal with feed-back. In spite of all these deficiencies, the fact that LMSs are Web based makes them very popular. Years ago many virtual study programs started by being text based, using HTML, PowerPoint, or PDF documents, eventually incorporating a wide range of multimedia technologies. Today animation and virtual reality (VR) are gaining space (Macromedia Flash, VRML and other animation and VR software) and the list of technologies used to design, develop and present e-learning applications is quite long: hypermedia, classroom response systems, blogs, e-mails, cooperative systems, computer aided assessment systems, electronic performance support systems, learning management systems screencasts, simulation, web 2.0 communities, ePortfolios, games, video and audio based courses, wiki, multimedia CD-ROMs and DVDs etc. In general, an e-learning application is using more than one of these techniques. We must also recognize thatWeb 2.0 has changed the whole pedagogical approach of learning. On institutional level we still have a hierarchical way of learning, with some collaborative learning elements introduced through technology. In parallel, a new collaborative learning environment, hundred per cent online, based on Web 2.0 services and technologies, has emerged. As a kind of response to the unsupervised learning possibilities, a special kind of software has appeared and was labelled *social software*.This kind of educational software is normally defined as web-based software programs that allow users to interact and share data with other users. It is a computer-mediated communication that has become very popular with social sites like MySpace and Facebook, media sites like Flickr and YouTube, and commercial sites like Amazon and eBay. From the technological point of view these programs share characteristics like open APIs, service oriented (customizable), and the ability to upload (data, media). Cooperative information sharing systems that enable collaborative work functions are named *collaborative software*.Having in mind that social aspects are important to be captured in the design of any e-learning application, a team of researchers from *Babes-Bolyai* University in Cluj-Napoca and *Lucian Blaga* University of Sibiu have started a large project DANTE - *Socio-Cultural Models implemented through multi-agent architecture for e-learning* - that has as main objective the development of a model for the virtual education system, student centered, that facilitates the learning through collaboration as a form of social interaction. The general architecture of the *e-Learning* proposed system is one with three levels (user, intermediary, supplier educational space), to each corresponding heterogeneous families of human agents and software (figure 2).

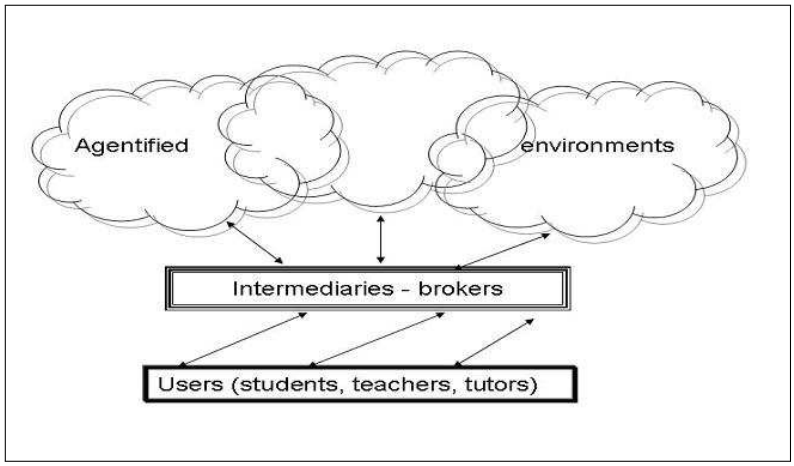


Figure 1: Three layered architecture of the DANTE system

In the virtual learning environment we have the corresponding agents. The human actors are interacting with the e-learning system via several agentified environments. The *teacher* (human agent) is assisted by two types of software agents: *personal assistant* (classic interface agent) and *didactic assistant*.The *SOCIAL agentified environment* has social agents and a database with *group models* (profiles of social behaviour).The *agentified DIDACTIC environment* assists the cognitive activities of the student and/or of the teachers. The student (human agent) evolves in an agentified environment with three types of agents. He/she has a personal assistant (software interface agent) who monitors all the student's actions and communicates (interacts) with all the other agents, with the agentified environments of other students and with the teacher's agentified environment. The student has at his/her disposal two more agents: the *TUTOR* and the *mediating agent*.The *TUTOR* assistant evaluates the educational objectives of the student and recommends her/him some kind of activities. The decisions are based on the knowledge of the students' cognitive and behaviour profiles. The *TUTOR* agent interacts with the personal assistant of the student, with the mediating agent and with the social agentified environment. As the system is conceived, the accent may

be put on collaboration activities between students, which consist in knowledge exchange, realization of common projects, tasks' negotiation, sharing resources, common effort for the understanding of a subject, problem-solving in-group. Several theories have inspired the models built in the DANTE system. They are briefly described in the following paragraphs.

Current conceptualizations of socio-cultural theory draw heavily on the work of Vygotsky (1986), as well as later theoreticians (see, for example, Wertsch, 1991, 1998). According to Tharp and Gallimore (1988) "This view [the socio-cultural perspective] has profound implications for teaching, schooling, and education. A key feature of this view of human development is that higher order functions develop out of social interaction. Vygotsky argues that a child's development cannot be understood by a study of the individual. We must also examine the external social world in which that individual life has developed. Through participation in activities that require cognitive and communicative functions, children are drawn into the use of these functions in ways that nurture and 'scaffold' them" (pp. 6-7). Kublin et al (1998) succinctly state that "Vygotsky (1934/1986) described learning as being embedded within social events and occurring as a child interacts with people, objects, and events in the environment" (p. 287). Vygotsky's theory of social cognitive development [16] is complementary to Bandura's social learning theory. Bandura's major premise is that we can learn by observing others. He considers vicarious experience to be the typical way that human beings change. He uses the term modelling to describe Campbell's two midrange processes of response acquisition (observation of another's response and modelling), and he claims that modelling can have as much impact as direct experience.

In modelling student's decisional behaviour towards educational objects, the **Fishbein's "reasoned action"** model of relationships among attitude, subjective norm, intention, and behaviour (Azjen and Fishbein, 1980) has been used, with a few changes. Fishbein's "reasoned action" model of relationships among attitude, subjective norm, intention, and behaviour (Azjen and Fishbein, 1980) and Fazio's (1986) attitude accessibility model were tested on a sample of 450 students. The results showed that the Fishbein model was more appropriate.

In order to describe differences in the way students learn, the concept of learning styles was used. Some students learn better reading by themselves the documentation, others use to ask questions. In the 70s, David **Kolb** was studying the Aristotelian learning by doing paradigm and developed a learning schema - **Learning Cycle** - consisting of four mandatory stages: real experience (concrete experience - **CE**), learning from experience by reflecting and observing (reflective observation - **RO**), abstract conceptualisation - **AC** - (identifying patterns, using theories and models, understanding what happened) and active experimentation - **AE**, trying out and planning for the next experience. Kolb defined four-type learning styles, each representing the combination of two preferred styles: **Diverging (CE/RO), Assimilating (AC/AE), Converging (AC/AE)**, and **Accommodating (CE/AE)**. Starting from Kolb's learning cycle, Peter Honey and Alan Mumford have shown ten years later that there are different learning styles and that, in general, a person is favouring only one way of learning. They have design a questionnaire (in fact there are two versions of the Learning Styles Questionnaire, the 80-item and the 40-item) to determine the preferred learning style. The **Honey** and **Mumford** four learning styles are:

- **Activists (Do)** - involving themselves fully in new experiences, open minded, enthusiastic, flexible, enjoying the here and now and being happy to be dominated by immediate experiences; acting first and considering consequences later; seeking activities to be centred around themselves.
- **Reflectors (Review)** - standing back and observing; reviewing the experience; collecting and analysing data about experience and events, slow to reach conclusions; maintaining a global view using information from past, present and immediate observation.
- **Theorists (Conclude)** - disciplined, aiming to fit things into rational order, adapting and integrating observations into coherent theories; thinking problems through in a vertical, step-by-step logical manner; attracted by systemic thinking, models, principles and theories.
- **Pragmatists (Plan)** - searching and new ideas and planning the next experiments; keen to put ideas, theories and techniques into practice; impatient with endless discussions.

The styles are not exclusive. A student can show different learning styles, but one is usually predominant. Learning styles as a description of the behaviours and attitudes, determine the preferred way to the students to learn.

Other learning styles have been defined for students in engineering and science starting from the observation that these students have an inductive way of learning (progressing from particulars - observations, measurements, and data- to generalities - governing rules, laws, and theories) in contrast with the teaching style that is deductive.

Analysing teaching practices and students' needs, **Felder** and **Silverman** [4] of North Carolina State University have built a model of learning styles for use by college instructors and students in engineering and sciences, a model that has subsequently been applied in a broad range of disciplines. They have designed an Index of Learning Styles -ILS to assess preferences on four scales: **sensing** (concrete, practical, oriented toward facts and procedures) **or intuitive** (discovering possibilities and relationships, disliking routine), **visual** (prefer visual representations of presented material, such as pictures, diagrams and flow charts) **or verbal** (prefer written and spoken explanations), **active** (learn by trying things out, enjoy working in groups) **or reflective** (thinking things through, prefer working alone or with one or two familiar partners), and **sequential** (linear thinking process, learn in incremental steps) **or global** (holistic thinking process, learn in large leaps). Each scale has 11 items.

In the followings a model for the student's behaviour is presented. The model is intended to be used by in the DANTE project by the student's virtual personal assistant and by the TUTOR agent.

2 The Student Model

The student (human agent) evolves in an agentified environment with three types of agents. He/she also has a personal assistant (software interface agent) who monitors all the students' actions and communicates (interacts) with all the other agents, with the agentified environments of other students and the TEACHER agentified environment. The student has at his/her disposal two more agents: *TUTOR* and the *mediating agent*.The TUTOR assistant evaluates the educational objectives of the student and recommends her/him some kind of activities. The decisions are based on the knowledge of the students' behavioural and cognitive profiles (which take into account the social component). The TUTOR agent interacts with the personal assistant of the student, with the mediating agent and with the social agentified environment. The mediating agent chooses an evolution mechanism of the solution to an exercise or a test proposed by the student, analyses the solution given by the student and produces feedback. The mediating agent can communicate with the personal assistants of other students. As the system is conceived, the accent is put on collaboration activities between students, which consist in knowledge exchange, realization of common projects, in groups, tasks' negotiation, resources' partition, common effort for the understanding of a subject, problem-solving in-group. The STUDENT model is a mixed one, embedding the behavioural theory with Vygotsky's socio-cultural theory and learning styles models.In our model it is considered that an action i is defined by a set of attributes A. An individual has to decide among a set I of n possible actions characterised by attributes and having a certain intensity i_k (the label of an action is its intensity):

$$I = \{i_k, k = 1,2,\ldots,n\} \text{ and } A_k = \{a_{ki} i = 1,2,\ldots,m\} \tag{1}$$

For example, when test results are less than the accepted one, the TUTOR is alerting the student and her/his agent STUDENT in order to fix the problem. Based on the student's profile, the TUTOR will recommend one action or a sequence of actions to be performed by the student (read study materials and enrolled for online self-tests, subscribe to a working group on the subject or to participate in a consultation session on a Saturday morning). For each action the STUDENT is *evaluating a quality-cost function* where tutor's satisfaction is opposed to real costs of the action (intellectual effort, time consumed, preferred timing). The evaluation of the quality cost-function is influenced also by factors inner to individuals. At individual level we are considering that the evaluation is also influenced by two categories of factors: ***beliefs*** (cognitive) and ***affects*** (emotive).

Beliefs are associations between actions and their attributes. Individual beliefs are cognitive; they depend on the level of education, culture and on the group's beliefs. The strength of the belief is directly determined by the strength of the association between an action and a certain attribute and in general does not depend on the true value of the association. They do not evaluate the *quality-cost function* at all; the belief is strong enough to be recorded in memory and to become automate. For example, in the former situation, some of the students can believe that the effort to acquire more knowledge is vain, too time consuming and too complicated especially in the time intervals between exam sessions. Even if the evaluation of the *quality-cost function* shows that more reading will enhance test results, the belief is stronger and they act consequently.)

Affects are feelings or desires associated with certain stimuli. There are many conceptual models of the affective component of behaviour. We will briefly remember some of these models. The *functional theory of attitude* considers that affects help individuals to accomplish a certain actions by application of prior knowledge, value expressions, and adjustments and by ego defence. According to the *Fishbein model* (Fig. 2) the decision to perform a certain action is directly influenced by the link between beliefs and affective responses. If the beliefs are strong and favourable for a certain action, the affective response is positive. This can be formalized as follows:

$$i_k = \sum_i B_{ki} E_{ki} \tag{2}$$

where

i_k is the intensity or the rank of an action., $k = 1, \ldots, n$;

B_{ki} is the belief that the action of intensity i_k posses the attribute a_{ki}, $i = 1, \ldots, m$

E_{ki} is the evaluation or utility (desirability) of attribute a_{ki}, $i = 1, \ldots, m$

At the end of the decisional process the student will perform the action with the maximum intensity i_k.

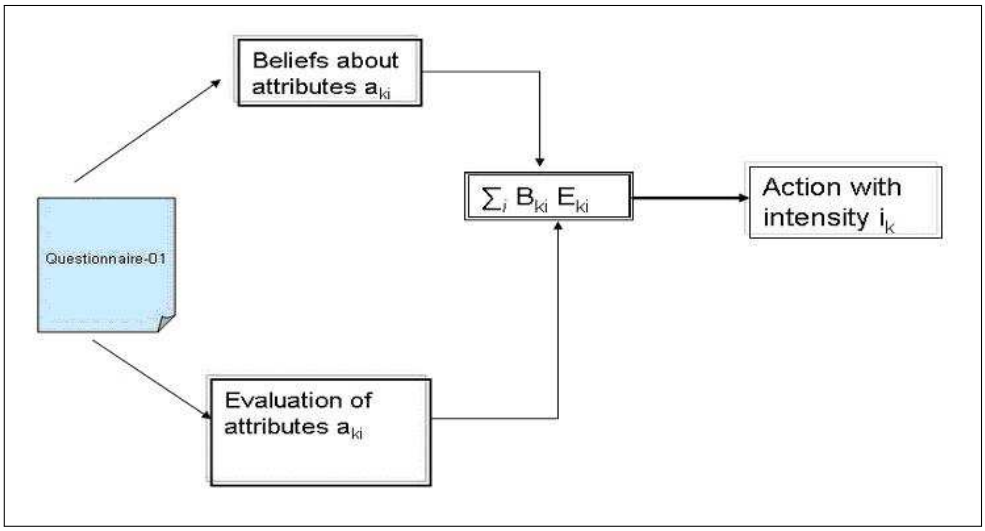


Figure 2: Fishbein model

The attributes of an action can be weighted in function of their importance for the task to be performed by the student and its personal preferences for performing other actions that are not linked with the educational process. For example, certain students will not participate in the consultation sessions that are taking place on Saturdays morning because they are involved every Saturday morning in sport activities and for them this is important. The beliefs about the existence of an attribute in an action, its desirability and importance are obtained from questionnaires addressed to students. For example, if an action (participating in project P1) is believed to have the attribute "high marks-60 points out of 100", the utility of this attribute is great, but if the action is believed to have also the attribute "high degree of difficulty" that has a low desirability, on the overall the action may have a lower intensity than another one (participate in another project P2) that has a different utility/desirability for its attributes.

The STUDENT personal assistant can use a modified Fishbein model where ranks are assigned to the attributes of an action and can be compared with those of other actions providing the same kind of results.

In the case of competing actions, the *belief-importance model* [5, 6, 7, 8] is used. In our study, 450 students were asked to indicate their belief toward educational objects (actions) and their decisional behaviour. The educational objects (actions) had prior validated attributes (same for all) and their importance. In order to express importance a constant sum scale was used; 100 importance points were distributed among the attributes based on their perceived relative importance. Data were obtained from previous surveys (312 students). Beliefs are rated from "excellent, very appealing" to "dreadful, not at all appealing" with scores from $+5$ to -5.

The belief-importance model is described by the next formula:

$$AT_k = \sum_i B_{ki} I_i;\ i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, m \tag{3}$$

where

AT_k is the attitude toward the action /object k;

B_{ki} is the belief that action k is desirable or not when comparing its attributes with those of competing actions;

I_i is the importance of the attribute i in selecting an action.

The belief-importance model is helping the investigator to have a better view of students' preferences and behaviours.

An important component of the attitude toward an action is ***intention***. Intention is a behavioural component. Behavioural intention describes the attitude not toward an action but toward performing an action. One of the models that take into account behavioural intention is the *theory of reasoned action* that states that behaviour is a direct result of intention and that there are involved two factors: *attitude toward an act and subjective norm*. The decident's attitude toward an act, α_k, is the sum of the decident's belief strength in the consequences resulting from performing a certain action (taking a certain decision) weighted by the evaluation of an anticipated outcome (positive benefit or avoidance of a negative consequence):

$$\alpha_k = \sum_i \beta_{ki} \varepsilon_{ki} \tag{4}$$

where

α_k is the attitude toward an action i_k, $k = 1, \ldots, n$;

β_{ki} is the belief that performing i_k will lead to an anticipated outcome i, $i = 1, \ldots, m$;

ε_{ki} is the evaluation or utility (desirability) of the outcome i, $i = 1, \ldots, m$.

The influence of the colleagues from the learning environment can be modelled by introducing the subjective norm. The ***subjective norm*** is the perception of an individual of what other people from the group think she/he should do with respect to certain behaviour, such as reading a specific article or enrolling for a pre-test. For sure our STUDENT is stressed about spending two more hours on reading if it is Saturday as this is a costly operation (being late for a date, etc.). But also the STUDENT knows that this is a special situation and that her boyfriend will appreciate a better mark to the exam. Normative beliefs and the motivation to comply with the beliefs are the two determinants of the subjective norm. This can be expressed as follows:

$$SN = \sum_i NB_{ki} MC_{ki} \tag{5}$$

where

SN is the subjective norm - the motivation toward an action i_k, $k = 1, \ldots, n$, as determined by the influence of the group;

NB_{kj} is the normative belief that people from the group (j) expect an individual to perform an action ik will lead to j, j=1,.., n;

MC_{kj} is the motivation to comply with the expectation of the group (j) j, $i = 1, \ldots, n$.

The theory of reasoned action is combining the attitude toward an act and the subjective norm:

$$DB = f[(BI) = f(\alpha_k) w_1 + (SN) w_2] \tag{6}$$

where

DB is the decisional behavior

BI is the behavioral intention

α_k is the attitude toward performing the action i_k

SN is the subjective norm

w_1 and w_2 are evaluation weights determined empirically

Student cognitive model is based on Vygotsky's theory. This theory of social cognitive development is basically considering that "social interaction plays a fundamental role in the development of cognition" (Kearsley 1994e). An important concept in Vygotsky's theory is that the potential for cognitive development is limited to a certain time span which he calls the "*zone of proximal development*" (Kearsley 1994e). He defines the zone of proximal development (ZPD) as having four learning stages. These stages "range between the lower limit of what the student knows and the upper limits of what the student has the potential of accomplishing" ([18]). The stages can be further broken down as follows: assistance provided by more capable others (coaches, experts, and teachers); self assistance; internalization automatization (fossilization); and de-automatization: (recursiveness through prior stages).

Vygotsky's theory also claims "that instruction is most efficient when students engage in activities within a supportive learning environment and when they receive appropriate guidance that is mediated by tools" (Vygotsky 1978, as cited in [18]). These instructional tools can be defined as "cognitive strategies, a mentor, peers, computers, printed materials, or any instrument that organizes and provides information for the learner." Their role is "to organize dynamic support to help [learners] complete a task near the upper end of their zone of proximal development [ZPD] and then to systematically withdraw this support as the [learner] move to higher levels of confidence." In our model these tools are represented by software agents.

Student population is considered a closed one and individuals are separated into groups, called *classes*. Students from a class communicate one with another and also with students from other classes. We will have intra-class and inter-class communication models and a different student-software agent communication model. A class consists of several teams.

3 Conclusions and future work

The socio-cultural model of the student as the main actor of a virtual learning environment is workable and appropriate to the DANTE requirements. The proposed model has been tested on a sample of 450 students from technical and scientific specialities. Adjustments have to be made periodically as we have observed changes in students' attitude toward educational objects and in learning styles. These changes are stronger as the students are approaching graduation, in the sense that they become more pragmatic. For example, the importance of the attribute "reward" for performing a certain educational task has a medium to minimum importance in the first years of study and is obtaining a maximum score on importance and on belief in the last two years of study.

As current intermediate results seem to show, the idea of modelling human and artificial agents using similar learning paradigms but avoiding dangerous equivalence is fruitful.

Future work is oriented toward the development of intra-class, inter-class communication models and student-software agent communication models.

References

[1] P. Honey, A. Mumford, (1982): Manual of Learning Styles. Honey. Maidenhead, 1982.

[2] P. Honey, and Mumford, Alan Using Your Learning Styles (Maidenhead: Peter Honey, 1995)

[3] D. A. KoIb, *Experiential Learning: Experience as the Source of Learning and Development*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

[4] R. M. Felder and L. K. Silverman, "*Learning Styles and Teaching Styles in Engineering Education*," Presented at the 1987 Annual Meeting of the American Institute of Chemical Engineers, New York, Nov. 1987.

[5] M. Fishbein, An Investigation of the Relationship between Beliefs About an Object and the Attitude Toward That Object, *Human Relations*, Vol. 16, pp.233-240, 1963.

[6] S. Shavitt, The Role of Attitude Objects in Attitude Functions, *Journal of Experimental Social Psychology*, Vol. 26, pp.124-148, 1990.

[7] J. van der Pligt, N. K. de Vries, Belief Importance in Expectancy-Value Models of Attitudes, *Journal of Applied Social Psychology* Vol. 28 (15) , 1339-1354, 1998.

[8] M. Conner, C. J. Armitage, Extending the Theory of Planned Behaviour: A Review and Avenues for Further Research, *Journal of Applied Social Psychology* Vol. 28 (15) , 1429-1464, 1998.

[9] Eric Abrahamson and Fombrun, J. Charles, Macrocultures: Determinants and Consequences. *Academy of Management Review*, 19:728-755, 1994.

[10] Thomas Burns and G. M. Stalker, *The Management of Innovation*. London: Tavistock, 1961.

[11] Jeffrey Pfeffer, Management as Symbolic Action: The Creation and Maintenance of Organizational *Paradigms, Research in Organizational Behavior*, 3: 1-52, 1981.

[12] T. Sawaragi, T. Ogura. Concept Sharing Between Human and Interface Agent Under Time Criticality. *Advances in networked enterprises. BASYS 2000, 4th IFIP/IEEE Int. Conf.* (L. Camarinha-Matos, H. Afsarmanesh, H-H. Erbe, Eds.), Kluwer Academic Publishers, Boston, 269-278, 2000.

[13] N. Balacheff, A modelling challenge: untangling learners' knowing. Journees Internationales d'Orsay sr les Sciences Cognitives: L'apprentissage, JIOSC2000, Paris, 2000.

[14] N. Balacheff, Teaching, an emergent property of eLearning environments. *IST 2000*, Nice, 2000

[15] j. D. Zapata-Rivera,; J. Greer, SMODEL Server: Student Modelling in Distributed Multi-Agent Tutoring *Systems. AI in Education, IOS Press*, 446-455, 2001.

[16] L. S. Vygotsky, Mind in Society: The development of Higher Psychological Processes, *Harvard University Press*, 1978.

[17] I. Moisil, C. Oprean, A. Nedan, From virtual to real systems - decisional behavioral patterns dynamics. A study based on the results of the e-castor control-panel simulation software, *WOSC'2005 Proceedings*, Maribor, Slovenia, 2005.

[18] B. Gillani, A. Relan, (1997). Incorporating Interactivity and Multimedia into Web_Based Instruction. In Khan, B. (Ed.), *Web-Based Instruction* (pp. 239-244). New Jersey: Educational Technology Publications, Inc.,pp. 239-244, 1997.

[19] R. Tharp, G. Gallimore, R. *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, England: Cambridge University Press, 1988.

Ioana I. Moisil
"Lucian Blaga" University of Sibiu, Romania
4, Emil Cioran Str., Sibiu 550025, Romania
E-mail: ioana.moisil@ulbsibiu.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 116-118

The Performance Optimization for Date Redistributing System in Computer Network

Parallel session invited paper

Grigor Moldovan, Mădălina Văleanu

Abstract: The paper treats of a problem related to the redistribution of distributed databases in a computer network. The minimal cost of the redistribution is established in the case of a distributed application.
Keywords: distributed database, redistribution

1 Introduction

Let us consider a database (tables), $b_i, i = \overline{1,n}$ distributed in r nodes belonging to a network of computer stations with own memory $S_i; i = \overline{1,r}$. Therefore:

B={$b_1, b_2, \ldots, b_n$}, S={$S_1, S_2, \ldots, S_r$}.

We shall identify the nodes and stations with memory supports S_i in S.

It is obvious that, at a certain moment, the stations of the computer network will provide a certain distribution, and grouping of the databases, respectively $b_i, i = \overline{1,n}$. If we take into account $n \geq r$ then, more b_i subbases will generally be present in a memory S_i.

To simplify, the study will consider that in each S_i station of the$r$ stations, we will have consecutively $d$taken subbases b_j, hence $n = d.r$.

A distributed application that requires programs running on the respective network leads to the accessing of the b_i subbases from the S_i nodes until the required result is reached.

Let us mark the data subbases that are accessed successively (some of them more times) in the vectorial form, as follows:

$B_L = (b_{m_1}, b_{m_2}, \ldots, b_{m_s})$
then
$L = (m_1, m_2, \ldots m_s); m_k \in$ {1, 2, ... ,n}

Remember that m_k identifies the place in the succession of accesses of the subbases in B performed.

In general, in the case of an access from S_i network node to a subbase found in S_j, a so-called penalty should also be considered (for instance: time, cost) noted with p_{ij} for all $i, j = \overline{1,r}$.

2 Grouping data (data subbases) on the network nodes

Let data (data subbases) be B={$b_1, b_2, \ldots, b_n$} ; their indices form the set I_n= {1,2, ... ,n}. The reorganisation of these data is defined by permutation indices

$$\sigma = \begin{pmatrix} 1\ 2\ \ldots k \ldots n \\ i_1 i_2 \ldots i_k \ldots i_n \end{pmatrix}$$

where $i_k \in I_n; k = \overline{1,n}$ are distinct. In fact, σ is a bijective application,
σ: {1,2, ... ,n} $\rightarrow$ {1,2, ... ,n} where $\sigma(k) = i_k$;
respectively $\sigma^{-1}(i_k) = k; k = \overline{1,n}$.

Permutation σ is also written as $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. If we have two permutations σ and τ , their produce $\sigma\tau$ is obtained by the composition of the two functions, so that

$\sigma\tau = (\sigma(\tau(1)), \sigma(\tau(2)), \ldots, \sigma(\tau(n)))$.

Let us mark with $Supp(\sigma)$, the permutation support σ , i.e. the set of the elements $i \in$ {1,2, ... ,n} having the property $\sigma(i) \neq j$.

A permutation σ is called cyclic of length *m, m≥ 2* if elements $i_1, i_2, \ldots, i_m \in Supp(\sigma)$ exist so as $\sigma(i_1) = i_2$, $\sigma(i_2) = i_3, \ldots \sigma(i_{m-1}) = i_m$, $\sigma(i_m) = i_1$.

It is known that any permutation that is different from the identical one can be written as a produce of distinct cycles.

Let us consider that $n = d.r$, i.e. on any S_i station in the *r*computer network we have *d* data subbases. Let us take the following reorganisation (distribution) of the *n* data subbases on the *r* workstations, successively:

$b_{\sigma^{-1}(1)}$	$b_{\sigma^{-1}(2)}$	$b_{\sigma^{-1}(d)}$	in S_1
$\ldots \ldots \ldots \ldots$			
$b_{\sigma^{-1}(1+(i-1)d)}$	$b_{\sigma^{-1}(2+(i-1)d)}$	$b_{\sigma^{-1}(id)}$	in S_i
$\ldots \ldots \ldots \ldots$			
$b_{\sigma^{-1}(1+(r-1)d)}$	$b_{\sigma^{-1}(2+(r-1)d)}$	$b_{\sigma^{-1}(rd)}$	in S_r

Example. Let *n=6, d=2, r=3,* hence *B = {*$b_1, b_2, \ldots, b_6$*} ; S ={*S_1, S_2, S_3*}* , hence, in each station there are two data subbases. Let us consider the permutation

$\sigma = \begin{pmatrix} 1\,2\,3\,4\,5\,6 \\ 4\,3\,1\,5\,2\,6 \end{pmatrix}$. Hence, $\sigma^{-1} = \begin{pmatrix} 1\,2\,3\,4\,5\,6 \\ 3\,5\,2\,1\,4\,6 \end{pmatrix}$.

The consecutive data subbases groups are: *{* b_3, b_5 *}* in S_1*; {*b_2, b_1*}* in S_2*; {*b_4, b_6*}* in S_3.

Let us note with S_i^σ the set of the indices of the subbases in the three S_i stations; that is

$S_1^\sigma = \{3,5\}$, $S_2^\sigma = \{2,1\}$, $S_3^\sigma = \{4,6\}$.
In the general case:

$$S_i^\sigma = \{\sigma^{-1}(1+(i-1)d), \sigma^{-1}(2+(i-1)d), \ldots, \sigma^{-1}(id)\}; i = \overline{1,r}.$$

Let us now note with $R_\sigma(k)$ the index of the station where the $b_k \in B$ subbase permutated with σ is. In the previous example b_4 is found in station S_3,hence $R_\sigma(4) = 3$, respectively $4 \in S_3^\sigma$.

Let us note with *I* the identical permutation, that is *I(k)=k; k=* $\overline{1,n}$. We find the property
$R_\sigma(k) = R_I(\sigma(k))$.

Trully
$R_\sigma(k) = i \Leftrightarrow k \in S_i^\sigma \Leftrightarrow \exists j, 1 \le j \le d : k = \sigma^{-1}(j+(i-1)d)$

Also,
$R_\sigma(k)=i \Leftrightarrow \exists$ j, 1≤ j ≤ d: σ(k) = j+(i-1)d ⇔ $\sigma(k) \in S_i^I$.

Consequently, the relation above occurs.

3 The cost of an application

Finding a result with the help of a distributed application *D*, managing the database
B={ $b_1, b_2, \ldots, b_n$*}* requires the consecutive use, once or more times, of some or all data subbases $b_i; i = \overline{1,n}$ established in the list

$B_L = (b_{m_1}, b_{m_2}, \ldots, b_{m_s}); m_k \in \{1,2,\ldots,n\}; k = \overline{1,s}$
Consider the sequence of successive accesses $L_D = (m_1, m_2, \ldots, m_s)$ of the subbases in *B* and the indices of the stations where the subbases of base *B* are. With these notations, we define the cost of a distributed application with the relationship

$$C(S,B,L_D,\sigma) = \sum_{k=1}^{s-1} (p_{R_\sigma(m_k),R_\sigma(m_{k+1})} + q_{R_\sigma(m_k)})$$

where $q_{R_\sigma}(m_k)$ represents the cost of the activities in station $S_{R_\sigma}(m_k)$.

Let us note with a_{ij} the number of the times $(m_k, m_{k+1}) = (i, j)$, $k = \overline{1, s-1}$;
$i,j \in \{1,2, \ldots ,n\}$, and with c_i , how many times
$m_k = i, k = \overline{1,s}; i, j \in \{1,2,\ldots,n\}$.
Then we can write

$$C(S,B,L_D,\sigma) = \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}p_{R_\sigma(m_k),R_\sigma(m_{k+1})} + c_i q_{R_\sigma}(m_k)) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}p_{R_I(\sigma(i)),R_I(\sigma(j))} + c_i q_{R_I}(\sigma(i))) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{\sigma^{-1}(i),\sigma^{-1}(j)}p_{R_I(i),R_I(j)} + c_i q_{R_I}(i))$$

If we note

$p_{R_I(i),R_I(j)} = p^*_{ij}; q_{R_I(i)} = q^*_i$

then the cost of a distributed application will be

$$C(S,B,L_D,\sigma) = \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{\sigma^{-1}(i),\sigma^{-1}(j)}p*_{ij} + c_i q^*_i)$$

4 Conclusions

1. In practice, we can consider that p_{ij} is symmetrical, hence p^*_{ij} will also be symmetrical, i.e. $p^*_{ij} = p^*_{ji}$. Penalties p_{ij} can be determined, for instance, with the help of statistical data after more runnings of the programs of the distributed application D.

2. If permutation σ is decomposed in a produce of cyclical permutations, the formula for the application cost can be simplified accordingly.

The fundamental problem, with respect to the distributed application D under consideration, consists in the determination of a permutation σ in the set of possible permutations P having elements *{1,2, ... , n},* indices of the B data subbases so that the cost of the use of the distributed application D programs would be minimal, i.e.
min { C(S,B,L_D,σ); $\sigma \in P$}.

It is obvious that the problem relates to combination, its solution is important when the distributed application D is used repeatedly. The problem is solved once and the advantage remains operational all along the use of the respective distributed application.

References

[1] L. Aspinall, *Data base. Re-organisation –Algorithms,* IBM, UKSC – 0029, 1972.

[2] G. Moldovan, I. Dzitac, *Sisteme distribuite - Modele matematice,* Ed. Univ. Agora, 2006.

[3] G. Moldovan, M. Valeanu, *Redistributing databases in a computer network,* Analele Univ. Bucureşti, Ser. Math.-Info., 56, 2006.

Grigor Moldovan
Babes-Bolyai University, Cluj-Napoca, Romania
Computer Systems Department
E-mail: moldovan@cs.ubbcluj.ro

Mădălina Văleanu
University of Medicine and Pharmacy "Iuliu Hatieganu", Cluj-Napoca, Romania
Medical Informatics and Biostatistics Department
E-mail: mvaleanu@umfcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 119-120

Natural Computing. Between Necessity and Fashion

Plenary invited paper

Gheorghe Păun

Extended Abstract

Many important steps in the history of computer science are related to and inspired from "computations" taking place in living cells and organisms. In the last decades this became a mainstream research direction – not to say a fashion, with important and well established areas, such as evolutionary computing and neural computing, and with exciting new areas, such as DNA and membrane (cellular) computing. Bio-inspired computing (hence in the benefit of computer science) has a counterpart in using computers (more general, computing theory) in biology, and this gave rise to several research directions, such as bio-informatics, system biology, computational biology, etc.

All these developments raise a plethora of questions, part of which will be touched in the talk. We only mention some of them here.

What is a computation? The standard answer is related to Turing machine – but can we go beyond Turing? Von Neumann has designed the first computers following Turing ideas (especially, the existence of universal Turing machines), but Turing-von Neumann computers have a series of drawbacks. Can these drawbacks be avoided? Can we imagine/realize computers with a basically different architecture? Can we find in biology suggestions for addressing the previous questions?

But, does nature compute? Using what data structures, operations with these data, kinds of "instructions" and of controls of them, kinds of "computer architectures"? Assuming as positive the answer to the question whether nature computes (and we can interpret as computations many processes taking place at the genomic, cellular, tissue – not to speak about the brain – level), can we learn something useful for traditional computer science or can we imagine a new kind of computers or new computing theories starting from biology? How all these can help in coping with computationally hard problems, hence in overpassing the limits of current computers?

The promises of natural computing are very high. Some directions of research try to improve the use of existing computers. This is the case of genetic algorithms (more generally, evolutionary computing), neural computing, swarm computing, etc. Other areas also promise new kinds of hardware – this is the case of DNA computing. Somewhere in the middle is membrane computing (P systems). Interestingly enough, Turing himself had proposals which can be considered now as pertaining to natural computing, but some of his papers in this area remained unpublished until recently.

In certain sense, the DNA molecule has an intrinsic computational universality, due to its organization (double stranded, based on nucleotides complementarity, etc.) How this power can be used – at a theoretical level and, hopefully, for practical computations? After the famous Adleman experiment (1994), DNA computing became sort of fashion, but the real life applications are still waited. Both theoretical developments are needed (e.g., related to space-time trade-off usually used in obtaining feasible solutions to computationally hard problems), but also bio-engineering breakthroughs. It was speculated that DNA behave better in its natural environment, the cell. This was one of the challenges of considering a computing model based on the cell biology, and in this way membrane computing was initiated.

All these should be put in relation with another fashionable area of research, called systems biology. After the genome successful project, the main challenge to computer science is the modeling and simulation of the living cell. However, using system theory in biology is an old research issue, which failed (in the times of Mesarović et al.) due to the lack of data and of computing power and resuscitated now, when such facilities are available.

Computing models inspired from biology in the aim of developing computer science prove to be now useful in this reverse direction, of modeling biological processes, with the hope of modeling the cell and of being useful to biology and bio-medicine.

Still, an important question remain: do we dream too much? In general, papers dealing with natural computing, system biology and related areas are optimistic and even over-optimistic. Indeed, biology proved to be a golden mine for computability and, moreover, it is not a good idea to underestimate the progresses in biology, computer science, etc. But, let us not forget that, besides technological difficulties, there are theoretical limits, such as the

(in)famous $\mathbf{P} \neq \mathbf{NP}$ problem/conjecture, Conrad impossibility theorems (programmability-universality, efficiency, and evolvability are three contradictory features of any computing model, no computing device can simultaneously have all these three good qualities...), Gandy theorems about the impossibility of computing beyond Turing as soon as four conditions are to be observed, and so on. Also, there are serious limits in what concerns the applications of computer science in biology; as Brooks suggested, "we might be missing something fundamental and currently unimagined in our models of biology".

Of course, these issues request a much longer/deeper discussion than the talk will make possible, that is why we mainly formulate them (as a challenge to the reader).

References

[1] L.M. Adleman: Molecular computation of solutions to combinatorial problems. *Science*, 226 (Nov. 1994), 1021–1024.

[2] R. Brooks: The relationship between matter and life. *Nature*, 409 (Jan. 2001), 409–411.

[3] M. Conrad: The price of programmability. In *The Universal Turing Machine: A Half–Century Survey* (R. Herken, ed.), Kammerer and Unverzagt, Hamburg, 1988, 285–307.

[4] R. Gandy: Church's thesis and principles for mechanisms. In *The Kleene Symposium* (J. Barwise et al., eds.), North–Holland, Amsterdam, 1980, 123–148.

[5] M.R. Garey, D.S. Johnson: *Computers and Intractability. A Guide to the Theory of NP–Completeness*. Freeman, San Francisco, 1979.

[6] S. Ji: The cell as the smallest DNA–based molecular computer. *BioSystems*, 52 (1999), 123–133.

[7] H. Kitano: Computational systems biology. *Nature*, 420 (Nov. 2002), 206–210.

[8] M.D. Mesarović: System theory and biology – view of a theoretician. In *System Theory and Biology* (M.D. Mesarović, ed.), Springer, New York, 1968, 59–87.

[9] Gh. Păun: Computing with membranes. *Journal of Computer and System Sciences*, 61, 1 (2000), 108–143.

[10] Gh. Păun: *Membrane Computing. An Introduction*. Springer, Berlin, 2002.

[11] G. Rozenberg, A. Salomaa: Watson–Crick complementarity, universal computations, and genetic engineering. *Techn. Report* 96–28, Department of Computer Science, Leiden Univ., Oct. 1996.

[12] C. Teuscher, ed.: *Alan Turing. Life and Legacy of a Great Thinker*. Springer, Berlin, 2003.

[13] O. Wolkenhauer: Systems biology: The reincarnation of systems theory applied in biology? *Briefings in Bioinformatics*, 2, 3 (2001), 258–270.

[14] The Web Page of Membrane Computing: `http://psystems.disco.unimib.it`

Gheorghe Păun
Institute of Mathematics of the Romanian Academy
PO Box 1-764, 014700 Bucureşti, Romania
E-mail: `george.paun@imar.ro, gpaun@us.es`

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 121-131

Fuzzy Set Theory in Boolean Frame

Workshop invited key lecture

Dragan G. Radojevic

Abstract: A Boolean frame consists of all axioms and theorems of Boolean algebra. Everything which satisfies all Boolean axioms and theorems is inside a Boolean frame and/or is an element of a corresponding Boolean algebra. Common to all conventional fuzzy set theories is the fact that they are not in the Boolean frame. Interpolative Boolean algebra (IBA) is a consistent real-valued ([0, 1]-valued) realization of finite (atomic) Boolean algebra. Since the axioms and laws of Boolean algebra are actually a matter of value independent structure of IBA elements, all axioms and all laws of Boolean algebra are preserved in any type of value realization (two-valued, three-valued,..., [0, 1]). To every element of IBA corresponds a generalized Boolean polynomial with the ability to process all values of primary variables from real unit interval [0, 1]. In this paper is given the theory of real sets (***R***-sets) based on IBA. ***R***-sets are generalized sets, as fuzzy sets. ***R***-sets are in the Boolean frame contrary to.
Keywords: Boolean algebra, Interpolative realization of Boolean algebra - IBA, Structure of Boolean algebra element, Principle of structural functionality, Generalized Boolean polynomial - GBP, Generalized product, Fuzzy sets, Real sets - ***R***-sets.

1 Introduction

Classical set theory (as classical logic, classical theory of relation generally) is in the Boolean frame and/or relies on classical two-valued realization of Boolean algebra. Two-valued realization of finite Boolean algebra is based on its homomorphic mapping on two-element Boolean algebra (truth, untruth in logic; belong, does not belong in theory of sets; in relation, not in relation in theory of relations etc.). A truth table is only a table representation of analyzed Boolean algebra homomorphic mapping on two-element Boolean algebra. All characteristics connected with a truth table such as the famous principle of truth functionality (and/or extensionality) are direct consequence of homomorphic mapping and hold only in a classical two-valued case.

In many real applications the classical two-valued ("black and white") realization of Boolean algebra [1] is not adequate. L. Zadeh, after his famous and distinguished contribution in the modern control theory, has ingeniously recognized the necessity of gradation in relations generally (theory of sets - fuzzy sets [2], logic - fuzzy logic [3], relations - fuzzy relations [4]).

Conventional fuzzy approaches rely on the same principle as many-valued (MV) logics [5]. MV-logics are similar to classical logic because they accept *the principle of truth-functionality* [5].A logic is truth functional if the truth value of a compound sentence depends only on the truth values of the constituent atomic sentences, not on their meaning or structure. The consequences of this direction are in the best way described by Lukasiewicz, the innovator of MV-logic:"*Logic* (truth functional) *changes from its very foundations if we assume that in addition to truth and falsehood there is also some third logical value or several such values,...*" [6].One "argument" for destroying the Boolean frame in treating gradation (MV-case) can be the definition of Boolean axioms of contradiction and excluded middle according to Aristotle: *The same thing cannot at the same time both belong and not belong to the same object and in the same respect* (Contradiction)... *Of any object, one thing must be either asserted or denied* (Excluded middle). If the goal is mathematics for gradation then it seems "reasonable" to leave these axioms as inadequate and accept the principle of truth functionality with all consequences or to go to the very source of Boolean algebra idea.

According to [7] fuzzy logic is based on truth functionality, since: "*This is very common and technically useful assumption*". A contrary example: "Our world is a flat plate" was also a very common and technically useful assumption in the Middle Ages!?

In the foundations of conventional MV approaches actually analyzed Boolean algebra is mapped onto a set of three or more scalars finally onto the real unit interval. **Sets of three or more scalars** $0, 1/2, 1, \ldots, 0, 1/n, \ldots, 1$, ... finally to the **real unit interval** $[0, 1]$ **are not Boolean algebras**. So, one cannot map Boolean algebra onto a set which is not Boolean algebra and in doing this preserve the properties of Boolean algebra and/or **no conventional fuzzy set theory** (fuzzy logic, theory of fuzzy relation) **is in the Boolean frame**. Structure $\langle [0, 1], T,S,N \rangle$, immanent

to fuzzy set theories, can't be Boolean structure, and the implementation of particular properties of Boolean algebra by creating corresponding T norms irresistibly resembles alchemy.

Two fuzzy sets are equal according to extensionality if and only if they have the same elements with equal membership functions. Extensionality is accepted as a fundamental principle in conventional fuzzy set theories from classical set theory. That extensionality is not a natural assumption in a fuzzy case, is illustrated by the following simple but representative enough example: Suppose a trivial set with only one element - a glass. Let a glass be half full with beer. According to extensionality, it follows that the set generated by property "beer" - half full glass with beer and set generated by property "no beer" - half empty glass, are equal (membership functions in both case have equal vale 0.5)! Actually these two sets haven't anything in common except the fact that they are in the same glass.

It is interesting that in his seminal paper [1] G. Boole has said: "...the symbols of the (logic) calculus do not depend for their interpretation upon the idea of quantity..." and only "in their particular application..., conduct us to the quantitative conditions of inference". So, according to G. Boole, the principle of truth functionality is not a fundamental principle (and as a consequence this principle can't be the basis of any generalization).

A very important question is: ***Can fuzziness and/or gradation be realized in a Boolean frame as a realization of Boolean algebra?*** We have obtained a ***positive answer*** to this question as an unexpected result, during solving the problem of fuzzy measure (or capacity) meaning in decision making by theory of capacity [8].

Boolean algebra (by axioms and theorems) defines value irrelevant properties which possess its elements. Classical two-valued realization of Boolean algebra (although extremely important) is only one of possible value realizations of Boolean algebra. The new approach to treating gradation in logic, theory of sets, relations etc., is based on interpolative realization of finite Boolean algebra (IBA) [9], [11].

IBA is real-valued and/or $[0,1]$-valued realization of finite Boolean algebra. In IBA to any element of analyzed finite Boolean algebra uniquely corresponds a Generalized Boolean polynomial (GBP) (as, for example any element of finite Boolean algebra uniquely corresponds to a disjunctive canonical form.). GBP processes values from real unit interval $[0,1]$. Values of GBP preserve partial order of corresponding Boolean algebra elements (based on value indifferent relation of inclusion) in all possible value realizations (so, for all values from real unit interval $[0,1]$) by relation $(\leq)$.

In case of $\boldsymbol{R}$-sets (idea of fuzzy sets realized in Boolean frame) properties are the domain of Boolean algebra. Any element of Boolean algebra represents a corresponding property, which generates a corresponding $\boldsymbol{R}$-set on the universe of discourses - value level.

Membership function of analyzed $\boldsymbol{R}$-set is defined by a corresponding GBP. Ω is a set of primary properties, whose characteristic is that no one can be represented as a Boolean function of the remaining primary properties. A power set of primary property set is a set $\alpha = P(\Omega)$ of atomic properties. Boolean algebra of properties BA(Ω) generated by the set of primary properties Ω is a power set of atomic properties BA(Ω) = P(α) and/or power set of power set of primary properties BA(Ω) = P(P(Ω)).GBP of any property - element of Boolean algebra is given by the superposition of relevant atomic GBP (as any element of finite Boolean algebra can be represented as a union of relevant atomic elements - a disjunctive canonical form). Any atomic element of analyzed Boolean algebra of properties generates, on the universe of discourses, a corresponding atomic $\boldsymbol{R}$-set. GBP of atomic property is a membership function of corresponding atomic $\boldsymbol{R}$-set. Class of atomic $\boldsymbol{R}$-sets is the partition of analyzed universe of discourses. Intersection of any two different atomic $\boldsymbol{R}$-sets is an empty set and union of all atomic $\boldsymbol{R}$-sets is equal to the universe of discourses. In a classical case any object of universe can be the element of only one atomic set, but in case of $\boldsymbol{R}$-sets it can be the element of a few atomic sets (in a special case, all of them) but so that the sum of values of corresponding atomic membership functions is equal to 1. Besides simultaneously owning a few atomic properties from the analyzed element of universe of discourses, intersection among atomic sets is always empty. So, simultaneously owning properties and the intersection of properties in the case of $\boldsymbol{R}$-sets are not synonyms as in a classical case. Simultaneous is only a necessary but not a sufficient condition for the intersection of two $\boldsymbol{R}$-sets in a general case. As a consequence, the known Aristotelian definition of excluded middle and contradiction are valid only for a classical case and can't be of any value for a general case and/or for $\boldsymbol{R}$-sets. Excluded middle and contradiction as well as all other axioms and theorems of Boolean algebra are value indifferent and they are valid in all possible realizations of $\boldsymbol{R}$-sets, as in the case of classical sets since classical sets can be treated as a special case of $\boldsymbol{R}$-sets. All results based on the theory of classical sets (finite number of classical sets) can be generalized straightaway by $\boldsymbol{R}$-sets. (For example: Kolmogorov theory of probability).

2 Real Sets (R-Sets) and Interpolative realization of Boolean algebra (IBA)

In a general case any element of ***R***-set, as of fuzzy set, has an analyzed property with *intensity* and/or *gradation*. ***R***-sets are generated by analyzed *proper properties* [1] on universe of discourses - value level. *Value indifferent (algebraic) characteristics* of proper properties are: *Boolean axioms and theorems*. The set of proper properties of interest BA_p is *Boolean algebra of analyzed properties*.

Set of primary properties: $\Omega a_1,...,a_n$, generates finite Boolean algebra of analyzed properties $BA_p(\Omega)$. The basic characteristic of any primary property $a \in \Omega$ is the fact that it can't be represented by Boolean function of the remaining primary properties. Boolean algebraic structure of analyzing proper properties is:

$$\langle BA_p(\Omega), \cap, \cup, C \rangle$$

Any element $\varphi \in BA - p(\Omega)$ of finite Boolean algebra of proper properties can be uniquely represented by the following disjunctive canonical form:

$$\varphi = \bigcup_{S \in P(\Omega) | \sigma_\varphi(S)=1} \alpha(S).$$

Where:

$\alpha(S) = \bigcap_{a_i \in S} a_i \bigcap_{a_j \in \Omega \setminus S} Ca_j, (S \in P(\Omega))$; *is atomic property*, atomic element of $BA_p(\Omega)$;

$\sigma_\varphi(S), (S \in P(\Omega))$; *is structure* of analyzed property $\varphi \in BA_p(\Omega)$.

Structure of any property $\varphi \in BA_p(\Omega)$ is given by the following expression:

$$\sigma_\varphi(S) = \begin{cases} 1, & \varphi \cap (S) = \alpha(S) \\ 0, & \varphi \cap (S) = \underline{0} \end{cases}$$

$(\underline{0}, \varphi \in BA(\Omega); S \in P(\Omega))$.

Fundamental structure's characteristic is ***principle of structural functionality*** [2]: *Structure of any combined property (element of Boolean algebra of analyzed properties) can be directly calculated on the base of structures of its components and the following rules*:

$$\begin{aligned} \sigma_{\varphi \cup \psi}(S) &= \sigma_\varphi(S) \vee \sigma_\psi(S), \\ \sigma_{\varphi \cap \psi}(S) &= \sigma_\varphi(S) \wedge \sigma_\psi(S), \\ \sigma_{C\varphi}(S) &= \neg \sigma_\varphi(S), \end{aligned}$$

$(S \in P(S))$.

Where: $\neg$ unary and $\vee, \wedge$ binary classical two-valued Boolean operators:

$$\begin{array}{cccccccc} \vee\rfloor & 0 & 1 & \vee\rfloor & 0 & 1 & & \lfloor\neg \\ 0 & 0 & 0; & 0 & 0 & 1; & 0 & 1\,. \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{array}$$

Structures of primary properties, given by the following set functions:

$$\sigma_{A_i}(S) = \begin{cases} 1, & A_i \in S \\ 0, & A_i \notin S \end{cases};$$

$$(S \in P(\Omega), A_i \in \Omega).$$

[1] or Boolean properties-unary relations

[2] Famous *principles of truth functionality* on value level is only the figure of irrelevant principle of structural functionality and is valid only for two-valued case

2.1 Generalized Boolean Polynomial

Generalized Boolean polynomials uniquely correspond to elements of Boolean algebra of analyzed properties - generators of ***R***-sets. ***R-characteristic function*** of any ***R***-set is its generalized Boolean polynomial.

Atomic R-set $A^{\otimes}(S)$ is value realization of corresponding atomic $\alpha(S),(S\in P(\Omega))$ property on analyzed universe of discourses **X**. ***R***-characteristic function $A^{\otimes}(S):X\to[0,1]$ of any atomic ***R***-set $A^{\otimes}(S),(S\in P(\Omega))$ is defined by corresponding **atomic generalized Boolean polynomial**, [8,9]:

$$A^{\otimes}(S)(x)=\sum_{K\in P(\Omega\setminus S)}(-1)^{|K|}\bigotimes_{A_i\in K\cup S}A_i(x)$$

$(A_i\in\Omega,x\in X)$.

Example: *Atomic Boolean polynomials - atomic* ***R****-characteristic function for the case when set of primary properties is* $\Omega=a,b$ *, are given in the following table*:

Table 1: *Example of Atomic Boolean polynomials.*

S	$A(S)$	$A^{\otimes}(S)(x)$
$\varnothing$	$A^c\cap B^c$	$1-A(x)-B(x)+A(x)\otimes B(x)$
A	$A\cap B^c$	$A(x)-A(x)\otimes B(x)$
B	$A^c\cap B$	$B(x)-A(x)\otimes B(x)$
A,B	$A\cap B$	$A(x)\otimes B(x)$

Any element of universe of discourses $x\in X$ has analyzed property $\varphi\in BA_p(\Omega)$ with intensity which is given by corresponding ***generalized Boolean polynomial***:

$$\begin{aligned}\varphi^{\otimes}(x) &= \sum_{S\in P(\Omega)|\sigma_\varphi(S)=1}A^{\otimes}(S)(x)\\ &= \sum_{S\in P(\Omega)}\sigma_\varphi(S)A^{\otimes}(S)(x),\end{aligned}$$

$(S\in P(\Omega),x\in X)$.

Generalized Boolean polynomial - ***R***-characteristic function, of $\varphi\in BA_p(\Omega)$ can be represented as scalar product of two vectors:

$$\varphi^{\otimes}(x)=\vec{\sigma}_\varphi(S)\vec{A}^{\otimes}(x).$$

Where

$$\begin{aligned}\vec{\sigma}_\varphi &= \lfloor\sigma^v_\varphi(S)|S\in P(\Omega)\rfloor,\\ \vec{A}^{\otimes}(x) &= [A^{\otimes}(S)(x)|S\in P(\Omega)]^T,(x\in X).\end{aligned}$$

are ***structure vector*** *of* φ and ***vector of atomic properties*** respectively. It is clear that the structure vector has algebraic nature, since it is value indifferent and preserves all axioms

$$\begin{array}{ll}\vec{\sigma}_{\varphi\cup(\psi\cup\phi)}=\vec{\sigma}_{(\varphi\cup\psi)\cup\phi} & \vec{\sigma}_{\varphi\cap(\psi\cap\phi)}=\vec{\sigma}_{(\varphi\cap\psi)\cap\phi}\\ \vec{\sigma}_{\varphi\cup\psi}=\vec{\sigma}_{\psi\cup\varphi} & \vec{\sigma}_{\varphi\cap\psi}=\vec{\sigma}_{\psi\cap\varphi}\\ \vec{\sigma}_{\varphi\cup(\varphi\cap\psi)}=\vec{\sigma}_\varphi & \vec{\sigma}_{\varphi\cap(\varphi\cup\psi)}=\vec{\sigma}_\varphi\\ \vec{\sigma}_{\varphi\cup(\psi\cap\phi)}=\vec{\sigma}_{(\varphi\cup\psi)\cap(\varphi\cup\phi)} & \vec{\sigma}_{\varphi\cap(\psi\cup\phi)}=\vec{\sigma}_{(\varphi\cap\psi)\cup(\varphi\cap\phi)}\\ \vec{\sigma}_{\varphi\cup C_\varphi}=\vec{1} & \vec{\sigma}_{\varphi\cap C_\varphi}=\vec{0}\end{array}$$

and theorems of Boolean algebra:

$$\begin{array}{ll}\vec{\sigma}_{\varphi\cup\varphi}=\vec{\sigma}_{\varphi} & \vec{\sigma}_{\varphi\cap\varphi}=\vec{\sigma}_{\varphi}\\ \vec{\sigma}_{\varphi\cup 0}=\vec{\sigma}_{\varphi} & \vec{\sigma}_{\varphi\cap 1}=\vec{\sigma}_{\varphi}\\ \vec{\sigma}_{\varphi\cup 1}=\vec{1} & \vec{\sigma}_{\varphi\cap 0}=\vec{0}\\ \vec{\sigma}_{C0}=\vec{1} & \vec{\sigma}_{C1}=\vec{0}\\ \vec{\sigma}_{C(\varphi\cup\psi)}=\vec{\varphi}_{C\varphi\cap C\psi} & \vec{\sigma}_{C(\varphi\cap\psi)}=\vec{\varphi}_{C\varphi\cup C\psi}\\ \vec{\varphi}_{CC\varphi}=\vec{\varphi}_{\varphi} & \end{array}$$

Any $\boldsymbol{R}$-set $\varphi^{\otimes}$, generated by corresponding property $\varphi\in BA_p(\Omega)$, can be represented as union of relevant atomic sets:

$$\varphi^{\otimes}=\bigcup_{S\in P(\Omega)|\sigma_{\varphi}(S)=1} A^{\otimes}(S).$$

Structures of proper properties preserve value irrelevant characteristics of $\boldsymbol{R}$-sets, actually their Boolean nature.

Generalized Product

In generalized Boolean polynomials there figure two standard arithmetic operators $+$ and $-$, and as a third *generalized product*$\otimes$. *Generalized product* is any function $\otimes:[0,1]\times[0,1]\to[0,1]$ that satisfies all four axioms of ***T-norms*** [12]:

Commutativity:

$$\begin{gathered}A(x)\otimes B(x)=B(x)\otimes A(x),\\ (A,B\in\Omega,\ A(x),B(x)\in[0,1],\ x\in X).\end{gathered}$$

Associativity:

$$\begin{gathered}A(x)\otimes B(x)\otimes C(x)=B(x)\otimes A(x)\otimes C(x),\\ (A,B,C\in\Omega,\ A(x),B(x),C(x)\in[0,1],\ x\in X).\end{gathered}$$

Monotonicity:

$$\begin{gathered}A(x)\le B(x)\Rightarrow A(x)\otimes C(x)\le B(x)\otimes C(x),\\ (A,B,C\in\Omega,\ A(x),B(x),C(x)\in[0,1],\ x\in X).\end{gathered}$$

Boundary:

$$\begin{gathered}A(x)\otimes 1=A(x),\\ (A\in BA(\Omega),\ A(x)\in[0,1],\ x\in X).\end{gathered}$$

and plus one additional axiom:

Non-negativity condition

$$\begin{gathered}\sum_{K\in P(\Omega\setminus S)}(-1)^{|K|}\bigotimes_{A_i\in K\cup S}A_i(x)\ge 0,\\ (\Omega=\{A_1,...,A_n\},S\in P(\Omega),A_i(x)\in[0,1],x\in X).\end{gathered}$$

Additional axiom "non-negativity" ensures that the values of atomic Boolean polynomials are non-negative: $A^{\otimes}(S)(x)\ge 0,(S\in P(\Omega),x\in X)$ As a consequence all elements of Boolean algebra and/or membership functions of all $\boldsymbol{R}$-sets are non-negative.

Comment: *Generalized product for* $\boldsymbol{R}$*-sets is just an arithmetic operator and as a consequence has a crucially different role from the role of the T-norm in conventional fuzzy set theories, where it is a set algebraic operator.*

***Example**: In the case $\Omega = \{a,b\}$generalized product, according to axioms of non-negativity can be in the following interval*[3] :

$$\max(a+b-1,0) \leq a \otimes b \leq \min(a,b).$$

Membership functions of intersection and union of two $\boldsymbol{R}$-sets are given by the following expressions:

$$(\varphi \cap \psi)^{\otimes}(x) = (\vec{\sigma}_{\varphi} \wedge \vec{\varphi}_{\psi})\vec{A}^{\otimes}(x),$$
$$(\varphi \cup \psi)^{\otimes}(x) = (\vec{\sigma}_{\varphi} \vee \vec{\varphi}_{\psi})\vec{A}^{\otimes}(x),$$
$$(\varphi, \psi \in BA(\Omega), x \in X)$$

Membership function of analyzed $\boldsymbol{R}$-set complement is:

$$(\varphi^{c})^{\otimes}(x) = 1 - \varphi^{\otimes}(x), (\varphi \in BA(\Omega), x \in X).$$

In new approach excluded middle and contradiction are always valid such as all other axioms and theorems of Boolean algebra for all possible generalized products:

$$\begin{aligned}
(\varphi \cup \varphi^{c})^{\otimes}(x) &= (\vec{\sigma}_{\varphi} \vee \vec{\sigma}_{\varphi})\vec{A}^{\otimes}(x), \\
&= \vec{1}\vec{A}^{\otimes}(x), \\
&= 1; \\
(\varphi \cap \varphi^{c})^{\otimes}(x) &= (\vec{\sigma}_{\varphi} \wedge \vec{\sigma}_{\varphi})\vec{A}^{\otimes}(x), \\
&= \vec{0}\vec{A}^{\otimes}(x), \\
&= 0; \\
(\varphi \in BA(\Omega), x \in X).
\end{aligned}$$

Example: The main characteristics of $\boldsymbol{R}$-sets are illustrated on the example of Boolean lattice of $\boldsymbol{R}$-sets generated by primary $\boldsymbol{R}$-sets $\Omega = \{A,B\}$represented in fig. 1.

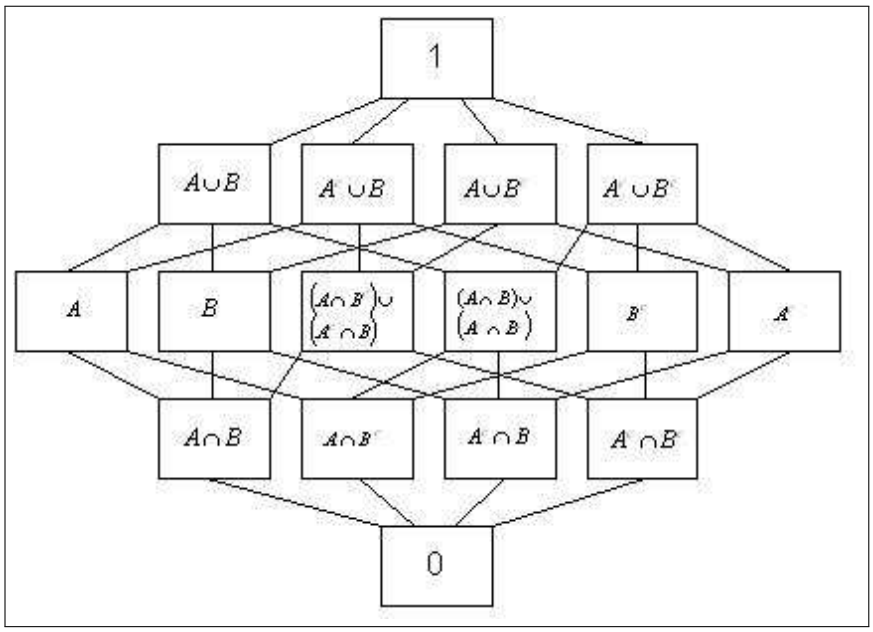


Figure 1: Boolean lattice generated by $\Omega = \{A,B\}$

The corresponding structure vectors
The generalized Boolean polynomials of corresponding sets are
Hasse diagram of corresponding generalized Boolean polynomials is represented in fig. 3.

[3] $max(a+b-1,0)$ is no more the low bound of feasible interval for generalized product in the case $|\Omega| \geq 3$.

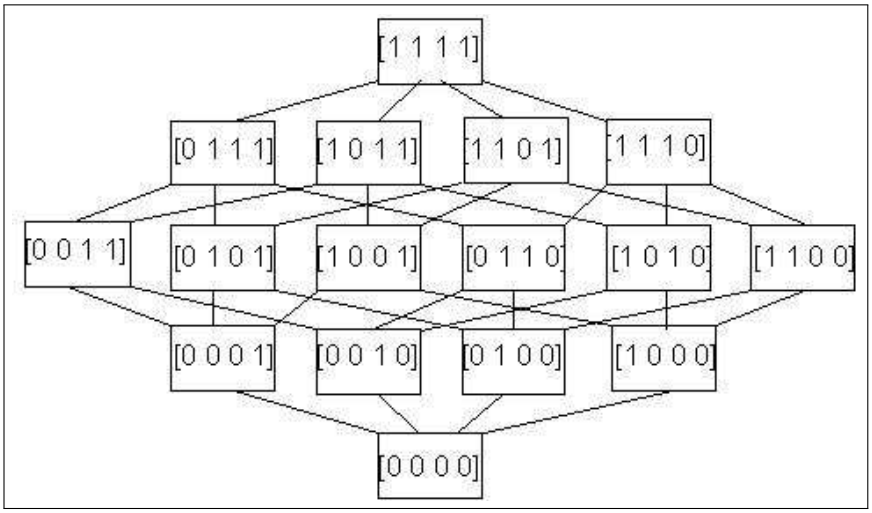


Figure 2: Boolean lattice of structure vectors

Table 2: *Example of Atomic Boolean polynomials.*

$(A\cap B)^{\otimes}(x) = A(x)\otimes B(x)$,
$(A\cap B^c)^{\otimes}(x) = A(x) - A(x)\otimes B(x)$,
$(A^c\cap B)^{\otimes}(x) = B(x) - A(x)\otimes B(x)$,
$(A^c\cap B^c)^{\otimes}(x) = 1 - A(x) - B(x) + A(x)\otimes B(x)$,
$A(x) = A(x)$,
$B(x) = B(x)$,
$((A\cap B)\cup(A^c\cap B^c))^{\otimes}(x) = 1 - A(x) - B(x) + 2A(x)\otimes B(x)$,
$((A\cap B^c)\cup(A^c\cap B))^{\otimes}(x) = A(x) + B(x) - 2A(x)\otimes B(x)$,
$(B^c)^{\otimes}(x) = 1 - B(x)$,
$(A^c)^{\otimes}(x) = 1 - A(x)$,
$(A\cup B)^{\otimes}(x) = A(x) + B(x) - A(x)\otimes B(x)$,
$(A^c\cup B)^{\otimes}(x) = 1 - A(x) + A(x)\otimes B(x)$,
$(A\cup B^c)^{\otimes}(x) = 1 - B(x) + A(x)\otimes B(x)$,
$(A^c\cup B^c)^{\otimes}(x) = 1 - A(x)\otimes B(x)$,
$(A\cap A^c)^{\otimes}(x) = 0$,
$(A\cup A^c)^{\otimes}(x) = 1$,
$(B\cap B^c)^{\otimes}(x) = 0$,
$(B\cup B^c)^{\otimes}(x) = 1$,

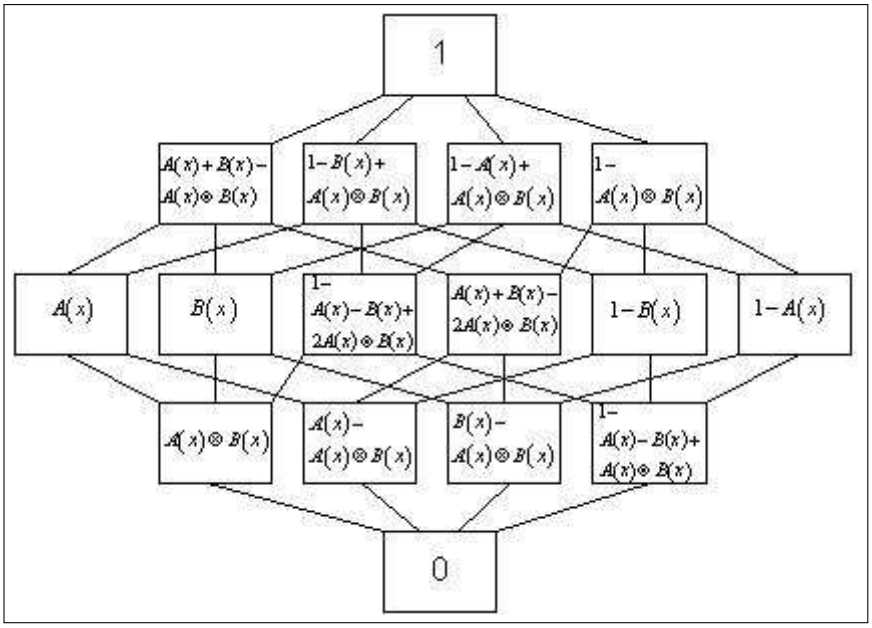


Figure 3: : Lattice of generalized Boolean polynomials generated by $\Omega = \{A, B\}$ for $x \in X$

Realization of all possible set functions for the given ***R***-sets A and B, in the case when the generalized product is given as ***min*** function is represented in fig. 4.

If we compare the example of ***R***-sets with corresponding classical sets given in fig 5:

it is clear that all properties of classical case are preserved one to one in generalized ***R***-sets case. Actually in the case of ***R***-sets interpretation is richer. In the case of classical sets one object of universe of discourses can be the element of only one atomic set, but in the case of ***R***-sets it can be the member of two and more atomic ***R***-sets but so that sum of corresponding membership function values is equal to 1.

2.2 R-partition

R-partition is consistent generalization of classical sets partition. Collection of atomic ***R***-sets $\{A^{\otimes}(S)|S \in P(\Omega)\}$ is R-partition of analyzed universe of discourses , since: (a) atomic sets are pair wise mutually exclusive:

$$A^{\otimes}(S_i) \cap A^{\otimes}(S^j) = \begin{cases} A^{\otimes}(S), & i = j \\ \varnothing, & i \neq j \end{cases};$$

$$\left((A^{\otimes}(S_i) \cap A^{\otimes}(S_j))(x) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} ; x \in X \right)$$

and

(b) they *cover* the universe X :

$$\bigcup_{S \in P(\Omega)} A^{\otimes}(S) = X; \left(\sum_{S \in P(\Omega)} A^{\otimes}(S)(x) = 1; (x \in X) \right).$$

Atomic ***R***-sets from the previous example are given in the following figure with GBP as their membership functions and their structures:

In the case of classical sets there is one additional constraint: any element of universe of discourses $x \in X$ belongs to only one classical atomic set $A(S), (S \in P(\Omega))$:

$$A(S_i)(x) = 1 \Rightarrow A(S_j)(x) = 0, S_j \neq S_i,$$
$$(S_i, S_j \in P(\Omega)).$$

This constraint is not general and it is not valid in general ***R***-sets case.

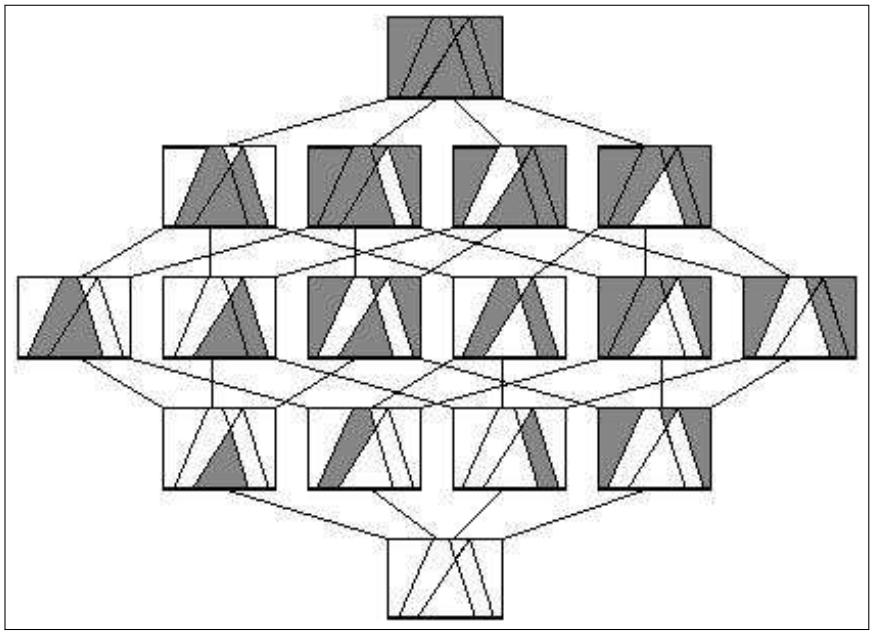


Figure 4: Hasse diagram of R-sets generated by $\Omega = \{A, B\}$ for $\otimes := min$

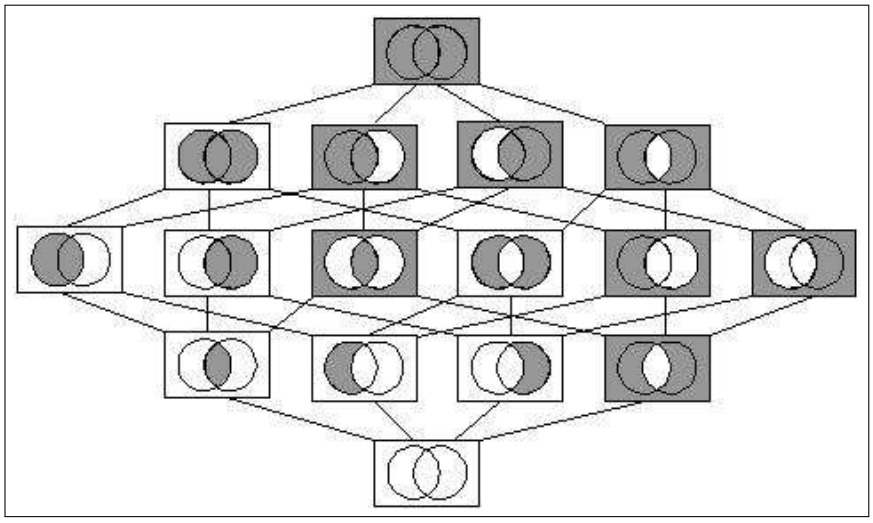


Figure 5: Hasse diagram of classical sets

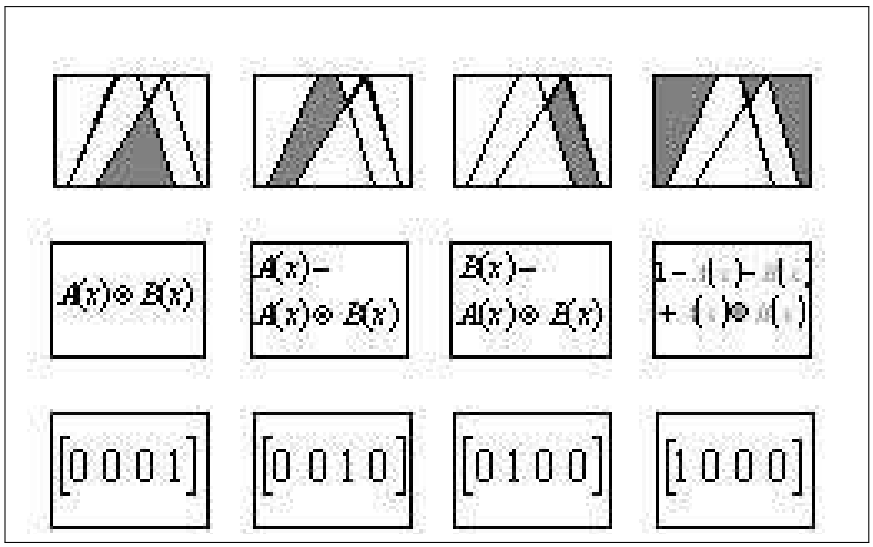


Figure 6: : Atomic $\boldsymbol{R}$-sets for $\otimes := min$

It is clear that all properties of the classical set algebra are preserved in the case of $\boldsymbol{R}$-sets algebra and/or by using $\boldsymbol{R}$-set approach (realization of IBA algebra) one can treat gradation in the Boolean frame, contrary to all fuzzy sets approaches.

3 Conclusions

Interpolative Boolean algebra (IBA) is a consistent real-valued and/or [0, 1]-valued realization of finite (atomic) Boolean algebra. In IBA to every element of any finite Boolean algebra, there corresponds a generalized Boolean polynomial with the ability to process all values of primary variables from real unit interval [0, 1]. Since the axioms and laws of Boolean algebra are actually a matter of value independent structure of IBA elements, all axioms and all laws of Boolean algebra are preserved in any type of value realization (two-valued, three-valued,..., [0, 1]).

Having recognized the constraints of classical mathematical approaches to real problems (for example: complete theory in the domain of automatic control is possible only for linear systems), L. Zadeh concluded that computation by words is very important since natural languages can be used for description of completely new results (for example quant physics) without changing its fundament. Objects of natural language are not as precise as mathematical objects; on the contrary, the number of words in a dictionary should be infinite, what is of course impossible. Non-preciseness immanent to a natural language is formalized by introducing gradation so an analyzed object can possess a property with some intensity or it can be a member of a generalized and/or fuzzy set. To the grammar of a natural language corresponds in mathematics algebra. Conventional approaches to theory of fuzzy sets are far from the original idea of computing with words since one natural language has one grammar but conventional approaches offer infinitely many algebras which are used depending on analyzed problem. One cannot imagine a poet who changes the grammar of a natural language dependent on a poem's contents!

The new approach offers as a solution the same algebra - Boolean algebra, both for the classical mathematical black and white view and for real applications with gradation. In the classical case one uses two-valued realization of Boolean algebra but in the generalized case one uses real-valued ([0, 1] - valued) realization of finite Boolean algebra. So, the new approach opens the door to the original Zadeh's idea of computing by words in one natural way. All results based on classical two-valued approaches by applying new approach can be straightaway generalized (theory of sets, logic, relations and their applications).

Possibilities of approaches based on IBA are illustrated on real sets (R-sets) as consistent realization of the idea of fuzzy sets - all laws of set algebra are preserved in the general case, contrary to conventional fuzzy approaches.

References

[1] R. G. Boole: The Calculus of Logic, *Cambridge and Dublin Mathematical Journal*, Vol. III, pp. 183-198, 1848.

[2] L. Zadeh: Fuzzy Sets, *Information and Control*, no. 8, pages 338-353. 1965.

[3] L. Zadeh: Bellman R.E., Local and fuzzy logics, *Modern Uses of Multiple-Valued Logic*, J.M. Dunn and G. Epstein (eds.), Dordrecht: D. Reidel, 103-165, 1977.

[4] L. Zadeh: Man and Computer, Bordeaux, France, 130-165, *Outline of a new approach to the analysis of complex systems and decision processe*s, IEEE 1972.

[5] S. Gottwald: *A Treats on Many-Valued Logics*, volume 9 of Studies in Logic and Computation. Research Studies Press, Bladock, 2000.

[6] J. Lukasiewicz: *Selected Works*. (ed.: L. Borkowski), North-Holland Publ. Comp., Amsterdam and PWN, Warsaw, 1970.

[7] P. Hajek: *Metamathematics of Fuzzy Logic, Trends in Logica* - Studia logica library, Kluwer Academic Publishers, Dodrecth /Boston/London, 1998.

[8] D. Radojevic: New [0,1]-valued logic: A natural generalization of Boolean logic, *Yugoslav Journal of Operational Research - YUJOR*, Belgrade, Vol. 10, No 2, 185-216, 2000.

[9] D. Radojevic: Interpolative relations and interpolative preference structures, *Yugoslav Journal of Operational Research - YUJOR* , Belgrade, Vol. 15, No 2, 2005.

[10] D. Radojevic: Logical measure - structure of logical formula, in *Technologies for Constructing Intelligent Systems 2*: Tools, Springer, pp 417-430, 2002.

[11] D. Radojevic: Interpolative realization of Boolean algebra as consistent frame for gradation and/or fuzziness, Studies in Fuzziness and Soft Computing: Forging New Frontiers: Fuzzy Pioneers II, Editors M. Nikravesh, J. Kacprzyk, L. Zadeh, Springer, pp. 295-318, 2007.

[12] R. Sikorski: *Boolean Algebras*, Springer-Verlag, Berlin, New York, 1964.

[13] E. P. Klement, Mesiar R., Pap E.: *Triangular Norms*, Kluwer Academic Publ, Dodrecht, 2000.

Dragan G. Radojevic
Mihajlo Pupin Institute
Volgina 15, 11000 Belgrade, Serbia
E-mail: dragan.radojevic@imp-automatika.bg.ac.yu

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 132-138

Intelligent Systems

Plenary invited paper & workshop invited key lecture

Imre J. Rudas, János Fodor

Abstract: In this paper we give an overview of intelligent systems. We discuss the notion itself, together with diverse features and constituents of it. We concentrate especially on computational intelligence and soft computing.
Keywords: intelligent systems, computational intelligence, soft computing.

1 Introduction

Intelligent systems (IS) provide a standardized methodological approach to solve important and fairly complex problems and obtain consistent and reliable results over time [2]. Extracting from diverse dictionaries, *intelligence* means the ability to comprehend; to understand and profit from experience. There are, of course, other meanings such as ability to acquire and retain knowledge; mental ability; the ability to respond quickly and successfully to a new situation; etc.

The definition of intelligent systems is a difficult problem and is subject to a great deal of debate. From the perspective of computation, the intelligence of a system can be characterized by its flexibility, adaptability, memory, learning, temporal dynamics, reasoning, and the ability to manage uncertain and imprecise information [9].

Independently from the definition, there is not much doubt that artificial intelligence (AI) is an essential basis for building intelligent systems. According to [13], AI consists of two main directions. One is *humanistic AI* (HAI) that studies machines that think and act like humans. The other one is *rationalistic AI* (RAI) that examines machines that can be built on the understanding of intelligent human behaviour. Here are some illustrative explanations from [9] where references to their original sources can also be found.

HAI is the art of creating machines that perform functions that require intelligence when performed by people. It is the study of how to make computers do things at which, at the moment, people are better. *RAI* is a field of study that seeks to explain and emulate intelligent behavior in terms of computational processes. It is the branch of computer science that is concerned with the automation of intelligent behavior.

Intelligent systems as seen nowadays have more to do with rationalistic than with humanistic AI. In addition to HAI features, IS admits intelligent behaviour as seen in nature as a whole; think, for example, on evolution, chaos, natural adaptation as intelligent behaviour. Moreover, IS are motivated by the need to solve complex problems with improving efficiencies.

Based on these and other similar considerations, an acceptable definition of intelligent systems was formulated in [9]. We adopt it here as follows.

Definition 1. [9] An *intelligent system* is a system that emulates some aspects of intelligence exhibited by nature. These include learning, adaptability, robustness across problem domains, improving efficiency (over time and/or space), information compression (data to knowledge), extrapolated reasoning.

The main aim of the present paper is to give an overview of diverse features and constituents of intelligent systems. After highlighting the notion of computational intelligence and its relationship to artificial intelligence, we go on with soft computing and hybrid systems.

2 Computational Intelligence

The development of digital computers made possible the invention of human engineered systems that show intelligent behaviour or features. The branch of knowledge and science that emerged together and from such systems is called *artificial intelligence*. Instead of using this general name to cover practically any approach to intelligent systems, the AI research community restricts its meaning to *symbolic representations and manipulations in a top-down way* [3]. In other words, AI builds up an intelligent system by studying first the structure of the problem (typically in formal logical terms), then formal reasoning procedures are applied within that structure.

Alternatively, non-symbolic and bottom-up approaches (in which the structure is discovered and resulted from an unordered source) to intelligent systems are also known. The conventional approaches for understanding and

predicting the behavior of such systems based on analytical techniques can prove to be inadequate, even at the initial stages of establishing an appropriate mathematical model. The computational environment used in such an analytical approach may be too categoric and inflexible in order to cope with the intricacy and the complexity of the real world industrial systems. It turns out that in dealing with such systems, one has to face a high degree of uncertainty and tolerate imprecision, and trying to increase precision can be very costly [11].

In the face of difficulties stated above fuzzy logic (FL), neural networks (NN) and evolutionary computation (EC) were integrated under the name *computational intelligence* (CI) as a hybrid system. Despite the relatively widespread use of the term CI, there is no commonly accepted definition of the term.

The birth of CI is attributed to the IEEE World Congress on Computational Intelligence in 1994 (Orlando, Florida). Since that time not only a great number of papers and scientific events have been dedicated to it, but numerous explanations of the term have been published. In order to have a brief outline of history of the term the founding and most interesting definitions will be summarized now.

The first one was proposed by Bezdek [1] as follows.

Definition 2. [1] A system is called *computationally intelligent* if it deals only with numerical (low-level) data, has a pattern recognition component, and does not use knowledge in the AI sense; and additionally, when it (begins to) exhibit (i) computational adaptivity; (ii) computational fault tolerance; (iii) speed approaching human-like turnaround, and (iv) error rates that approximate human performance.

Notice how the role of pattern recognition is emphasized here. In addition, remark that Bezdek concerns an artificially intelligent system as a CI system whose "added value comes from incorporating knowledge in a nonnumerical way."

In [10], one of the pioneering publications on computational intelligence, Marks defined CI by listing the building blocks being neural nets, genetic algorithms, fuzzy systems, evolutionary programming, and artificial life. Note that in more recent terminology genetic algorithms and evolutionary programming are called by the common name evolutionary computing.

In the book [5] Eberhart *et al.* formulated their own definition and its relationship to the one of Bezdek.

Definition 3. [5] Computational intelligence is defined as a methodology involving computing (whether with a computer, wetware, etc.) that exhibits an ability to learn and/or deal with new situations such that the system is perceived to possess one or more attributes of reason, such as generalisation, discovery, association, and abstraction.

Eberhart *et al.* stress adaptation rather than pattern recognition (Bezdek). They say it explicitly: *computational intelligence and adaptation are synonymous.* That is, in this sense CI do not rely on explicit human knowledge [7]. Notice that adaptability is one of the key features of intelligent systems also in Definition 1.

Closing this section, we briefly recall three typical opinions on the relationship between AI and CI, leaving to the reader to judge them.

In [10] Marks wrote: "Although seeking similar goals, CI has emerged as a sovereign field whose research community is virtually distinct from AI." This opinion declares that CI means an alternative to AI.

Bezdek in [1], after an analysis based on three levels of system complexity, came up with the conclusion that CI is a subset of AI. This viewpoint was criticized in [5].

Fogel formulated a third opinion in [7]. Starting from *adaptation* as the key feature of intelligence, and observing that traditional symbolic AI systems do not adapt to new problems in new ways, he declares that AI systems emphasize *artificial* and not the *intelligence*. Thus, it may be inferred that AI systems are not intelligent, while CI systems are.

3 Soft Computing

Prof. Lotfi A. Zadeh [14] proposed a new approach for Machine Intelligence, separating Hard Computing techniques based Artificial Intelligence from Soft Computing techniques based Computational Intelligence (Figure 1).

Hard computing is oriented towards the analysis and design of physical processes and systems, and has the characteristics precision, formality, categoricity. It is based on binary logic, crisp systems, numerical analysis, probability theory, differential equations, functional analysis, mathematical programming, approximation theory and crisp software.

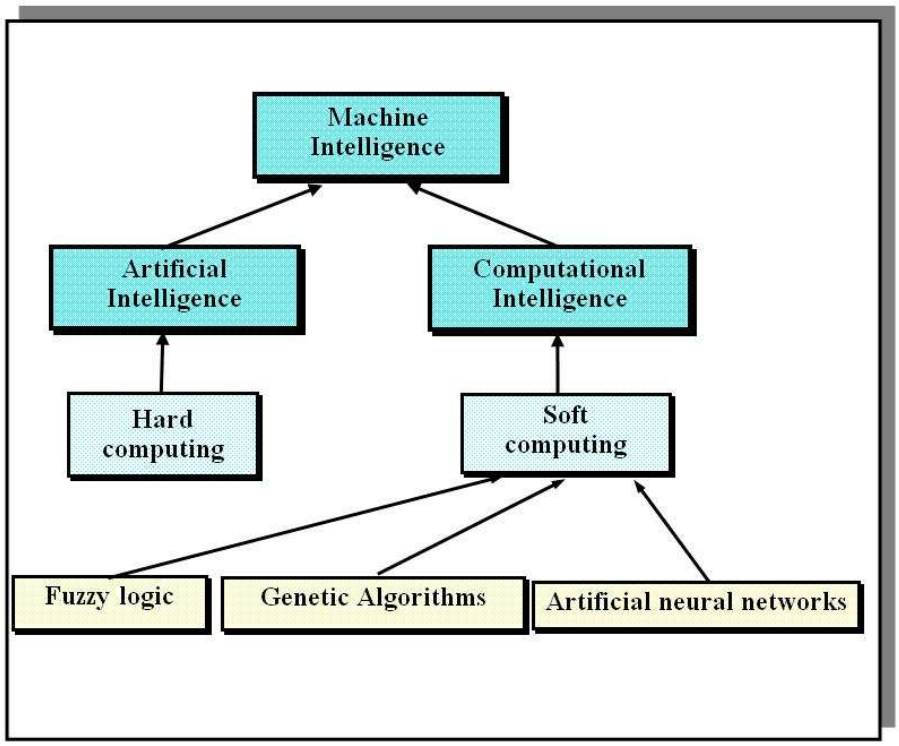


Figure 1: Artificial Intelligence vs. Computational Intelligence

Soft computing is oriented towards the analysis and design of intelligent systems. It is based on fuzzy logic, artificial neural networks and probabilistic reasoning including genetic algorithms, chaos theory and parts of machine learning and has the attributes of approximation and dispositionality.

Although in hard computing imprecision and uncertainty are undesirable properties, in soft computing the tolerance for imprecision and uncertainty is exploited to achieve an acceptable solution at a low cost, tractability, high Machine Intelligence Quotient (MIQ). Prof. Zadeh argues that soft computing, rather than hard computing, should be viewed as the foundation of real machine intelligence.

Soft computing, as he explains, is

- a consortium of methodologies providing a foundation for the conception and design of intelligent systems,
- aimed at a formalization of the remarkable human ability to make rational decision in an uncertain and imprecise environment.

The guiding principle of soft computing is: *Exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness, low solution cost and better rapport with reality.*

Fuzzy logic (FL) is mainly concerned with imprecision and approximate reasoning, neural networks (NN) mainly with learning and curve fitting, evolutionary computation (EC) with searching and optimization. The constituents of soft computing are complementary rather than competitive.

The experiences gained over the past decade have indicated that it can be more effective to use them in a combined manner, rather than exclusively. For example an integration of fuzzy logic and neural nets has already become quite popular (neuro-fuzzy control) with many diverse applications, ranging from chemical process control to consumer goods.

The constituents and the characteristics of hard and soft computing are summarized in Table I.

3.1 Fuzzy Logic

Fuzziness refers to nonstatistical imprecision and vagueness in information and data. Most concepts dealt with or described in our world are fuzzy. In classical logic, known as crisp logic, an element either is or is not a member

of a set. That is, each element has a membership degree of either 1 or 0 in the set. In a fuzzy set, fuzzy membership values reflect the membership grades of the elements in the set. Membership function is the basic idea in fuzzy set theory based on fuzzy logic, which is the logic of "approximate reasoning." It is a generalization of conventional (two-valued, or crisp) logic. Fuzzy sets model the properties of imprecision, approximation, or vagueness. Fuzzy logic solves problems where crisp logic would fail.

Fuzzy logic is being applied in a wide range of applications in engineering areas ranging from robotics and control to architecture and environmental engineering. Other areas of application include medicine, management, decision analysis, and computer science. New applications appear almost daily. Two of the major application areas are *fuzzy control* and *fuzzy expert systems.*

3.2 Neural Networks

An artificial neural network (briefly: neural network) is an analysis paradigm that is roughly modeled after the massively parallel structure of the brain. It simulates a highly interconnected, parallel computational structure with many relatively simple individual processing elements. It is known for its ability to deal with noisy and variable information.

There are five areas where neural networks are best applicable [6]:

- Classification;
- Content Addressable Memory or Associative Memory;
- Clustering or Compression;
- Generation of Sequences or Patterns;
- Control Systems.

3.3 Evolutionary Computing

Evolutionary computing comprises machine learning optimization and classification paradigms roughly based on mechanisms of evolution such as biological genetics and natural selection. The evolutionary computation field includes genetic algorithms, evolutionary programming, genetic programming, evolution strategies, and particle swarm optimization. It is known for its generality and robustness. Genetic algorithms are search algorithms that incorporate natural evolution mechanisms, including crossover, mutation, and survival of the fittest. They are used for optimization and for classification. Evolutionary programming algorithms are similar to genetic algorithms, but do not incorporate crossover. Rather, they rely on survival of the fittest and mutation. Evolution strategies are similar to genetic algorithms but use recombination to exchange information between population members instead of crossover, and often use a different type of mutation as well. Genetic programming is a methodology used to evolve computer programs. The structures being manipulated are usually hierarchical tree structures. Particle swarm optimization flies potential solutions, called particles, through the problem space. The particles are accelerated toward selected points in the problem space where previous fitness values have been high.

Evolutionary algorithms have been applied in *optimization* to multiple-fault diagnosis, robot track determination, schedule optimization, conformal analysis of DNA, load distribution by an electric utility, neural network explanation facilities, and product ingredient mix optimization. *Classification* applications include rule-based machine learning systems and classifier systems for high-level semantic networks. An application area of both optimization and classification is the evolution of neural networks.

4 Hybrid Systems

Hybrid systems combine two or more individual technologies (fuzzy logic, neural networks and genetic algorithms) for building intelligent systems. The individual technologies represent the various aspects of human intelligence that are necessary for enhancing performance. However, all individual technologies have their constraints and limitations. Having the possibility to put two or more of them together in a hybrid system increases the system's capabilities and performance, and also leads to a better understanding of human cognition.

Table I. The constituents and the characteristics of hard and soft computing after [11].

HARD COMPUTING		SOFT COMPUTING	
Based on	***Has the characteristics***	***Based on***	***Has the characteristics***
• binary logic • crisp systems • numerical analysis • differential equations • functional analysis • mathematical programming • approximation theory • crisp software	• quantitative • precision • formality • categoricity	• fuzzy logic • neurocomputing • genetic algorithms • probabilistic reasoning ▪ machine learning ▪ chaos theory ▪ evidental reasoning ▪ belief networks	• qualitative • dispositionality • approximation

Several models are used for integrating intelligent systems. The one used in [8] classifies hybrid architectures into the following four categories:

Combination: typical hybrid architecture of this kind is a sequential combination of neural networks and rule- or fuzzy rule-based systems.

Integration: this architecture usually uses three or more individual technologies and introduces some hierarchy among the individual subsystems. For example, one subsystem may be dominant and may distribute tasks to other subsystems.

Fusion: a tight-coupling and merging architecture, usually based on the strong mathematical optimization capability of genetic algorithms and neural networks. When other techniques incorporate these features, the learning efficiency of the resulting system is increased.

Association: the architecture that includes different individual technologies, interchanging knowledge and facts on a pairwise basis.

Lotfi A. Zadeh expectation was explained as follows: *"in coming years, hybrid systems are likely to emerge as a dominant form of intelligent systems. The ubiquity of hybrid systems is likely to have a profound impact on the ways in which man-made systems are designed, manufactured, deployed and interacted with."*

Fuzzy logic is mainly concerned with imprecision and approximate reasoning, neural networks mainly with learning and curve fitting, evolutionary computation with searching and optimization. Table II gives a comparison of their capabilities in different application areas, together with those of control theory and artificial intelligence.

Table II. after [12]

	Mathematical Model	Learning Data	Operator Knowledge	Real Time	Knowledge Representation	Nonlinearity	Optimization
Control Theory	*Good*	*X*	*Needs*	*Good*	*X*	*X*	*X*
Neural Network	*X*	*Good*	*X*	*Good*	*X*	*Good*	*Fair*
Fuzzy Logic	*Fair*	*X*	*Good*	*Good*	*Needs*	*Good*	*X*
Artificial Intelligence	*Needs*	*X*	*Good*	*X*	*Good*	*Needs*	*X*
Genetic Algorithms	*X*	*Good*	*X*	*Needs*	*X*	*Good*	*Good*

Explanation of Symbols: Good=Good or suitable, Fair=Fair, Needs=Needs some other knowledge or techniques, X=Unsuitable or does not require.

5 Summary and Conclusions

In this paper we gave an overview of intelligent systems, computational intelligence and soft computing. The notions have been discussed in considerable details. Essential features were highlighted together with typical applicational areas. In our plenary lecture at ICCCC 2008 we will analyse further the constituents and their simultaneous usage.

References

[1] J.C. Bezdek, "What is computational intelligence?", In: J.M. Zurada, R.J. Marks II, and C.J. Robinson, Eds., *Computational Intelligence, Imitating Life,* IEEE Computer Society Press, pp. 1-12, 1994.

[2] T.A. Byrd and R.D. Hauser, "Expert systems in production and operations management: research directions in assessing overall impact," *Int. J. Prod. Res.*, Vol. 29, pp. 2471-2482, 1991.

[3] B.G.W. Craenen, A.E. Eiben, "Computational Intelligence," *Encyclopedia of Life Support System,* EOLSS Co. Ltd., http://www.eolss.net, 2003.

[4] W. Duch, "What is Computational Intelligence and where is it going?", In: W. Duch and J. Mandziuk, Eds., *Challenges for Computational Intelligence,* Springer Studies in Computational Intelligence, Vol. 63, pp. 1-13, 2007.

[5] R. Eberhart, P. Simpson, and R. Dobbins, *Computational Intelligence PC Tools,* Academic Press, Boston, 1996.

[6] R.C. Eberhart and Y. Shui, *Computational Intelligence - Concepts to Implementations,* Elsevier, 2007.

[7] D. Fogel, "Review of computational intelligence imitating life,"*IEEE Transactions on Neural Networks,* Vol. 6, pp. 1562-1565, 1995.

[8] M. Funabashi, A. Maeda, Y. Morooka and K. Mori, "Fuzzy and neural hybrid expert systems: synergetic AI," *IEEE Expert*, August, pp. 32-40, 1995.

[9] K. Krishnakumar, "Intelligent systems for aerospace engineering – an overview," *NASA Technical Report*, Document ID: 20030105746, 2003.

[10] R. Marks, "Computational versus artificial," *IEEE Transactions on Neural Networks,* Vol. 4, pp. 737-739, 1993.

[11] I.J. Rudas, Hybrid Systems (Integration of Neural Networks, Fuzzy Logic, Expert Systems, and Genetic Algorithms), In: *Encyclopedia of Information Systems,* Academic Press, pp. 114-1 - 114-8, 2002.

[12] I.J. Rudas and M.O. Kaynak, "Techniques in Soft Computing and their Utilization in Mechatronic Products," In: C.T. Leondes, Ed., *Diagnostic, Reliability and Control System Techniques,* Gordon and Beach International Series in Engineering, Technology and Applied Science Volumes on Mechatronics Systems Techniques and Applications, Vol. 5, Singapore, pp. 101-138, 2000.

[13] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach,* Prentice Hall, New Jersey, 1995.

[14] L.A. Zadeh, "Fuzzy Logic and Soft Computing: Issues, Contentions and Perspectives," In: *Proc. of IIZUKA'94: Third Int.Conf. on Fuzzy Logic, Neural Nets and Soft Computing,* Iizuka, Japan, pp. 1-2, 1994.

[15] L.A. Zadeh, "Fuzzy Logic, Neural Networks, and Soft Computing," *Communications of the ACM,* Vol. 37, pp. 77-84, 1994.

Imre J. Rudas, János Fodor
Budapest Tech
Institute of Intelligent Engineering Systems
Bécsi út 96/b, H-1034 Budapest, Hungary
E-mail: {rudas,fodor}@bmf.hu

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 139-143

Mind Your Words! You Might Convey What You Wouldn't Like To

Workshop invited key lecture

Dan Tufiş

Abstract: Natural language ambiguity is a well-known challenge for the NLP community in deepening the performances of the computer programs when dealing with human languages. Language ambiguity as produced by humans, is often unnoticed and as such, is most of the times involuntary. However, in many cases ambiguity is purposely used for various reasons. In an original context, a sentence might be very clear with respect to the producer's intentions, but if it contains some unnoticed ambiguities (obliterated by the context), when put in another context, might convey a very different meaning, sometimes funny, sometimes embarrassing. We describe a system which is able to pinpoint such potential misinterpretations and advise the writer to make other lexical choices for his/her wording.

1 Introduction

Sentiment analysis has recently emerged as a very promising research area with multiple applications in processing arbitrary collections of text. The numerous related workshops (at least a dozen in the last 3-4 years, with the most recent one to be held in Marrakech associated with LREC 2008), as well as a two-weeks Summer School (Eurolan 2007) entirely dedicated to this topic, are obvious signs of the generalized interest in subjectivity analysis and opinion mining.

Sentiment can be expressed about works of art and literature, about the state of financial markets, about liking and disliking individuals, organizations, ideologies, and consumer goods. In making everyday decisions or in expressing their opinions on what they are interested in, more and more people are interacting on the so-called social web, reading others' opinions or sharing their experiences or sentiments on a wide spectrum of topics. The review sites, forums, discussion groups and blogs became very popular and opinions expressed therein are getting significant influence on people's daily decisions (buying products or services, going to a movie/show, traveling somewhere, forming an opinion on political topics or on various events, etc.). Decision makers, at any level, cannot ignore the "word-of-mouth" as sometimes the social web is dubbed. Research in the area of opinion finding and sentiment analysis is motivated by the desire to provide tools and support for information analysts in government, commercial, and political domains, who want to automatically track attitudes and feelings in the news and on-line forums [1].

There are various ways to model the processes of opinion mining and opinion classifications and different granularities at which these models are defined (documents vs. sentences). For instance, in reviews classification one would try to assess the overall sentiment of an opinion holder with respect to a product (positive, negative and possibly neutral). However, the document level sentiment classification is too coarse for most applications and therefore the most advanced opinion miners are considering the sentence level. At the sentence level, the typical tasks include identifying the opinionated sentences and the opinion holder, deciding whether these opinions are related or not to a topic of interest and classifications according to their polarity (positive, negative, undecided) and force.

Irrespective of the methods and algorithms (which are still in their infancy) used in subjectivity analysis, they exploit the pre-classified words and phrases as opinion or sentiment bearing lexical units. Such lexical units (also called senti-words, polar-words) are manually specified, extracted from corpora or marked-up in the lexicons such as General Inquirer (www.wjh.harvard.edu/ inquirer/homecat.htm), SentiWordNet [2] etc.

While opinionated status of a sentence is less controversial, its polarity might be rather problematic. The issue is generated by the fact that words are polysemous and the polarity of many senti-words depend on context (some time on local context some time on global context). Apparently, bringing into discussion the notion of word sense (as SentiWordNet does) solves the problem but this is not so. We argued elsewhere [3] that it is necessary to

make a distinction between words intrinsically bearing a specific subjectivity/polarity and the words the polarity of which should be relationally considered. The latter case refers to the head-modifier relations; compare the different polarities of the modifier *long* in the two contexts: "the response time is long" vs. "the engine life is long".

Research in this area has been for some time monolingual, focused on English, which is "mainly explained by the availability of resources for subjectivity analysis, such as lexicons and manually labeled corpora" [4]. Yet, in the recent years, there are more and more languages (including Romanian [4], [5]) for which required resources are developed, essentially by exploiting parallel data and multilingual lexicons. With more than 40 monolingual wordnets (see http://www.globalwordnet.org/), most of them aligned to the Princeton WordNet [6], the recent release of SentiWordNet, and several public domain language independent tools for opinion mining and sentiment analysis, the multilingual research in opinion mining and sentiment analysis has been boosted and more and more sophisticated multilingual applications are expected in the immediate future. Such an application is briefly introduced in the next section.

2 Analysis of potential unwanted connotations

Many commercials make clever use of the language ambiguity (e.g. puns, surprising word associations, images pushing-up a desired interpretation context etc.) in promoting various products or services. Many of these short sentences when used in regular texts might have their connotations obliterated by the context and unnoticed by the standard reader. This observation is also valid the other way around: specific sentences, conceived in the context of a given text, when taken out of their initial context and placed in a conveniently chosen new context may convey a completely new (potentially unwanted) message/attitude. It is relatively easy to find, especially in argumentative texts, examples of sentences which taken out of their context and maliciously used could have an adverse interpretation compared to the intended one.

The subjectivity and sentiment analysis methods are usually concerned with detecting whether a sentence is subjective and in case it is, establishing its polarity. Our approach takes a different position: we are interested in whether a given sentence, taken out of context, may have different subjective interpretations. We estimate, on a [0,1] scale, the potential of a sentence being objective (O), positively subjective (P) or negatively subjective (N), based on the senti-words in the respective sentence. Usually, these scores are uneven with one of them prevailing. We found that sentences which may have comparable subjective (positive, negative or both) and objective scores are easier to use in a denotation/connotation shift game.

The SentiWordnet [2] has been initially developed for English but the subjectivity information can be imported into any other language's wordnet which is aligned with Princeton WordNet or use it as an interlingual index. Such a wordnet, with subjectivity annotation imported from English (via the synset translation equivalence relations) will be referred to as a sentiwordnet.

3 CONAN (CONotation ANalyzer)

The CONAN system has been developed in a language independent way, and it should work for various languages, provided the analyzed texts are appropriately pre-processed and there are sentiwordnets available for the considered languages.

The necessary text preprocessing, required by CONAN includes: tokenization, tagging, lemmatization and chunking and, optionally, dependency linking. These fundamental operations for any meaningful natural language processing application have been largely described in previous publications and recently have been turned into public web-services [7] on our web server (https://nlp.racai.ro). Currently our linguistic web-services are available for Romanian and English.

Figure 1 exemplifies the result of preprocessing a Romanian sentence as mentioned above. The example

sentence (*Marele Juriu din Fulton a spus vineri că o investigare a alegerilor nu a produs nicio dovadă că ar fi avut loc nereguli.*) is extracted from the Romanian translation of the SemCor.

```
<s id="br-a01.1.1.ro">
    <w lemma="mare" ana="1+,Afpmsry" chunk="Np#1,Ap#1">Marele</w>
    <w lemma="juriu" ana="1+,Ncms-n" chunk="Np#1">Juriu</w>
    <w lemma="din" ana="5+,Spsa" chunk="Pp#1">din</w>
    <w lemma="Fulton" ana="8+,Np" chunk="Pp#1,Np#2"
                                    wns="ili:ENG20-00026769-n">Fulton</w>
    <w lemma="avea" ana="3+,Va--3s" chunk="Vp#1">a</w>
    <w lemma="spune" ana="1+,Vmp--sm" chunk="Vp#1,Np#3,Ap#2"
                                    wns="ili:ENG20-00976600-v">spus</w>
    <w lemma="vineri" ana="1+,Ncf--n" chunk="Np#3"
                                    wns="ili:ENG20-14305402-n">vineri</w>
    <w lemma="că" ana="31+,Csssp">că</w>
    <w lemma="un" ana="21+,Tifsr" chunk="Np#4">o</w>
    <w lemma="investigare" ana="1+,Ncfsrn" chunk="Np#4">investigare</w>
    <w lemma="al" ana="21+,Tsfs" chunk="Np#4">a</w>
    <w lemma="alegere" ana="1+,Ncfpoy" chunk="Np#4"
                                    wns="ili:ENG20-00171672-n">alegerilor</w>
    <w lemma="recent" ana="1+,Afpfp-n" chunk="Np#4,Ap#3">recente</w>
    <w lemma="nu" ana="7+,Qz" chunk="Vp#2">nu</w>
    <w lemma="avea" ana="3+,Va--3s" chunk="Vp#2">a</w>
    <w lemma="produce" ana="1+,Vmp--sm" chunk="Vp#2,Ap#4"
                                    wns="ili:ENG20-02079175-v">produs</w>
    <c>"</c><w lemma="nici_un" ana="22+,Dz3fsr---e" chunk="Np#5">nicio</w>
    <w lemma="dovadă" ana="1+,Ncfsrn" chunk="Np#5"
                                    wns="ili:ENG20-05487494-n">dovadă;</w>
    <c>"</c>
    <w lemma="că" ana="31+,Csssp">că</w>
    <w lemma="avea" ana="3+,Va--3" chunk="Vp#3">ar</w>
    <w lemma="fi" ana="3+,Vanp" chunk="Vp#3">fi</w>
    <w lemma="avea_loc" ana="1+,Vmp--sm" chunk="Vp#3,Np#6,Ap#5">avut_loc</w>
    <w lemma="neregulă" ana="1+,Ncfp-n" chunk="Np#6">nereguli</w>
    <c>.</c></s>
```

Figure 1: The preprocessing results of a sentence

After the text is preprocessed as required, the second phase identifies all senti-words, i.e. those words which in the support sentiwordnet (in our case the Romanian one) have at least one possible subjective interpretation (that is, their objectivity score is less than 1, see [8] for details). There has been mentioned by various authors [8, 9, 10] that the bag-of-words (BoW) approaches to subjectivity analysis is not appropriate since the subjectivity priors (the lexicon mark-up subjectivity) may be changed in context by the so-called valence shifters: intensifiers, diminishers and negations. The first two operators increase and respectively decrease the subjectivity scores (both the negative and the positive ones) while the latter complements the subjective values. As the valency shifter do not necessary act on the senti-word in their immediate proximity, the chunking pre-processing step mentioned earlier is necessary for taking care of delimiting the scope of the operators action. For instance in the sentence "He is NOT *clever* ENOUGH", the word in italics (clever) is a (positive) senti-word, while the upper case words are valence shifters: NOT is a negation and ENOUGH is a diminisher. The diminisher acts on the senti-word, while the negation act on the result of the diminisher: NOT(ENOUGH(clever)). As a consequence, the sentence above has a negative subjectivity score. In [3] we showed that some wrong subjectivity mark-up existing in SentiWordNet can be explained due to a BoW approach to sense definitions analysis. The majority of synsets with wrong computed subjectivity markup have in their definitions valence shifters which apparently were ignored.

The algorithm computes the subjectivity scores of the grammatical chunks (taking into account the subjectivity operators and their scopes) summing their values to get the score for the entire sentence. As shown in the snapshot in Figure 2, the user can ask for the most *objective* interpretation (the considered senses for the senti-words are the ones with the highest objective scores), the most *positive subjective* interpretation (the considered senses for

the senti-words are the ones with the highest subjective scores), the most *negative subjective* interpretation (the considered senses for the senti-words are the ones with the highest negative scores) or all the three interpretations (as shown in Figure 2).

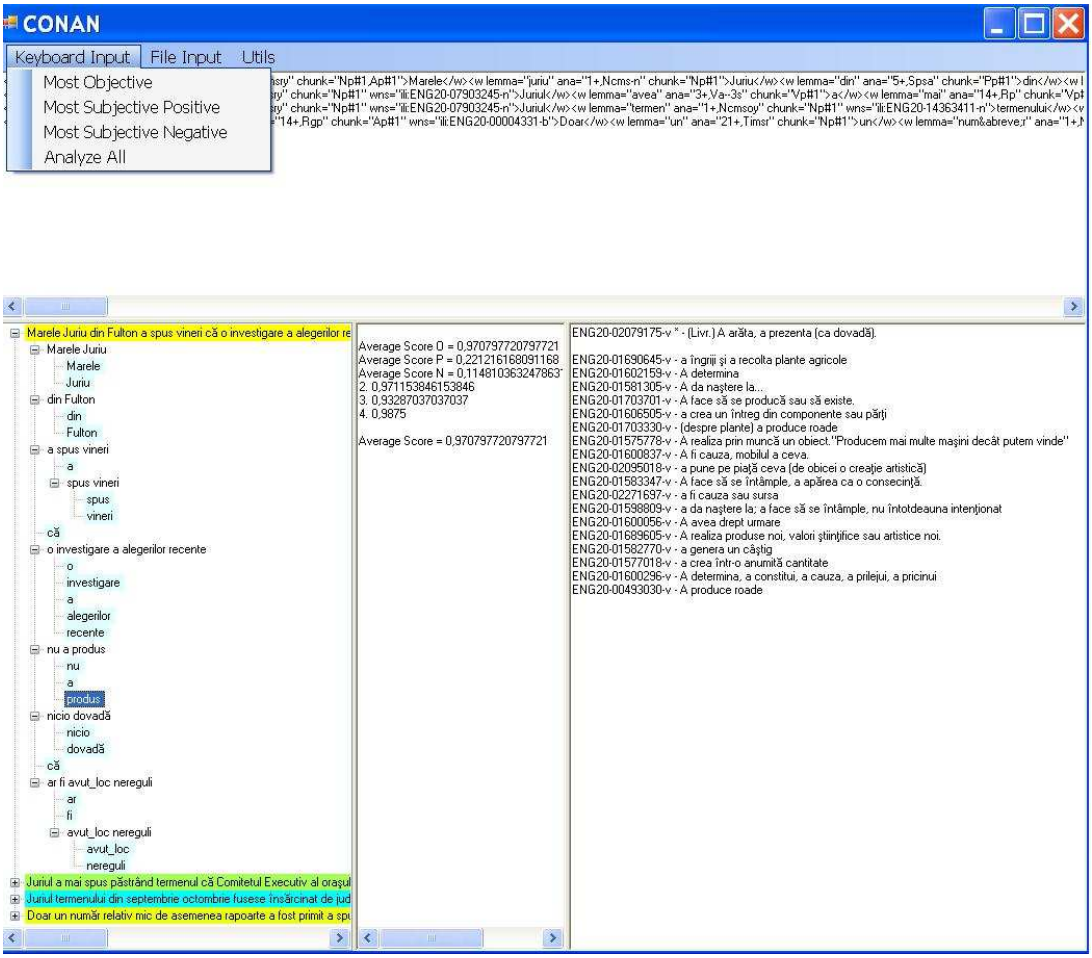


Figure 2: The CONAN interface

As mentioned in the previous section, for those sentences which get a most objective score close to 0.5, it is possible to have them interpreted subjectively (the O, P, N scores observe the relation O=1-(P+N), see [2]). Similarly, those subjective sentences which have comparable values for the P and N scores are potential trouble makers. The senti-words responsible for the possible interpretation shifts are highlighted by the interface and, via the wordnet support, the analysts may chose different words with less (or more) subjectivity load. The sentence exemplified in Figure 2, has a very high objective score (0,97) which make it an undisputable objective utterance.

References

[1] Wiebe Janyce, Wilson Theresa, Cardie Claire. Annotating Expressions and Emotions in Language. In Language, *Resources and Evaluation*, vol. 39, No. 2/3, pp. 164-210, 2005.

[2] Esuli Andrea, Sebastiani Fabrizio. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06)*, Genoa, Italy, pp. 417-422, 2006.

[3] Tufiş Dan. Subjectivity mark-up in WordNet: does it work cross-lingually? A case study on Romanian Wordnet. *Invited talk on the Panel "Wordnet Relations" at the Global WordNet Conference*, January 22-25, 2008.

[4] Tufiş Dan, Ion Radu. Cross lingual and cross cultural textual encoding of opinions and sentiments. *Invited talk to Eurolan Summer School on "Semantics, Opinion and Sentiment in Text"*, July 23-August 3, 2007, Iaşi.

[5] Mihalcea Rada, Banea Carmen, Wiebe Janyce. Learning Multilingual Subjective Language via Cross-Lingual Projections. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic, June, pp. 976-983, 2007.

[6] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

[7] Tufiş Dan, Ion Radu, Ceauşu Alexandru, Ştefănescu Dan. RACAI's Linguistic Web Services. *In Proceedings of 6th Conference on Language Resources and Evaluation (LREC-08)*, Marrakech, Marocco, 2008.

[8] Tufiş Dan, Ion Radu, Bozianu Luigi, Ceauşu Alexandru, Ştefănescu Dan. Romanian WordNet: Current State, New Applications and Prospects. In Attila Tanacs, Dora Csendes, Veronika Vincze, Christiane Fellbaum, Piek Vossen (eds.): *Proceedings of 4th Global WordNet Conference, GWC-2008*, University of Szeged, Hungary, January 22-25, pp. 441-452, 2008.

[9] Polanyi Livia, Zaenen Annie. Contextual Valence Shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Application*. Springer Verlag, 2006.

[10] Andreevskaia Alina, Bergler Sabine. Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses. *In Proceedings EACL-06 the 11rd Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, pp. 209-216, 2006.

[11] Miyoshi Tetsuya, Nakagami Yu. Sentiment classification of customer reviews on electric products. *IEEE International Conference on Systems, Man and Cybernetics*, pp. 2028-2033, 2007.

Dan Tufiş
Research Institute for Artificial Intelligence
Romanian Academy
13, 13 Septembrie, 050711, Bucharest
E-mail: tufis@racai.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 144-149

Statistical Edge Detectors Applied to SAR Images

Mohamed Airouche, Mimoun Zelmat, Madjid Kidouche

Abstract: Segmentation is a crucial step in automatic interpretation of images. However, the presence of speckle in Synthetic Aperture Radar (SAR) images makes their segmentation very difficult. Several methods of segmentation have been developed specifically to improve the interpretation of SAR images. In this paper, we present edge detectors applied to SAR images. First, we present the detector Ratio of Averages (ROA), which is based on the normalized ratio computation of intensity. But, this detector is optimum in the case of mono-edge. Second, we present the detector Ratio of Exponentially Weighted Averages (ROEWA), which is optimized for multi-edge model, in the Minimum Mean Square Error (MMSE) sense. Finally, we give experimental results of these operators on optical image, simulated SAR image and on ERS-1 SAR image.
Keywords: Edge Detection, Image Segmentation, Ratio of Average.

1 Introduction

Image segmentation is a fundamental step in image analysis. It consists of partitioning an image into homogeneous regions that share some common properties. There are two main approaches in image segmentation: edge- and region- based. Edge-based segmentation looks for discontinuities in the intensity of an image. Region-based segmentation looks for uniformity within a sub-region, based on a desired property, e.g. intensity, color, and texture. In this paper, segmentation by edge detection applied to ERS-1 SAR images is described. Edge detection plays a key role in image analysis. In images with no texture, an edge can be defined as the boundary between two regions with relatively distinct properties. The usual gradient-based edge detectors, developed for optical images, compute the difference between the local mean values on opposite sides of considered pixel.

In the last decade, Synthetic Aperture Radar (SAR) imaging systems have been widely used in the observation of the earth's surface. The main advantages of these systems are the ability to operate at any time of day, in any weather conditions, and to improve the image resolution for a given aperture size. The structures in SAR images give important contextual information useful to the detection and the classification of entities, as vegetation, urban area, and industrial area. However, because of the coherent nature of wave in radar, SAR images are degraded by specific noise called "speckle". The presence of speckle in SAR images makes their segmentation very difficult. In SAR images, the usual gradient-based edge detectors give more false edges in areas of high reflectivity than in areas of low reflectivity [2, 3]. Then, the presence of speckle in SAR images makes the usual edge detectors inefficient and complicates edge detection.

Several edge detectors have been developed specifically for SAR images [1, 2, 4, 5]. In this paper, we present the Ratio of Averages (ROA) operator based on the normalized ratio computation of local averages on opposite sides of the central pixel, for different directions and sizes of the analyzing window. However, this operator uses arithmetic means computed on opposite halves of the analyzing window. In this case, we consider only one edge is present in the analyzing window. This is only optimal in the idealized mono-edge case. The Ratio of Exponentially Weighted Averages (ROEWA) is an edge detector optimized for stochastic multi-edge model. It consists basically on the normalized ratio computation of means on opposite sides of the central pixel, with non-uniform weighting. The coefficients weighting of the pixels is considered as a function of the distance to the central pixel. In this case, several edges are considered present in the analyzing window.

2 Ratio of Averages (ROA) operator

The Ratio of averages (ROA) is an edge detector with Constant False Alarm Rate (CFAR), developed specifically for SAR images. In this operator, an analyzing window of size $N \times N$ is considered. This analyzing window is divided into two zones relative to central pixel. It calculates the ratio of arithmetic means $\bar{I}_1, \bar{I}_2$ of intensity on opposite halves of the analyzing window

$$r = \frac{\bar{I}_1}{\bar{I}_2} \tag{1}$$

On the analyzing window, we consider the presence of one edge only. Figure 2 illustrates the idealized mono-edge model in the horizontal direction. To make the operator independent of the scanning direction the ratio can be normalized. Both cases are present. The first one is based on the choice of the maximum ratio, than we obtain a ratio of intensity r greater than one. But, in this case the variation interval of the normalized ratio is not defined. The second one is based on the choice of the minimum ratio, this gives a limited interval. In this paper we consider the second case, and the ratio is normalized to lie between zero and one

$$r = \min\left\{\frac{\bar{I}_1}{\bar{I}_2}, \frac{\bar{I}_2}{\bar{I}_1}\right\} \tag{2}$$

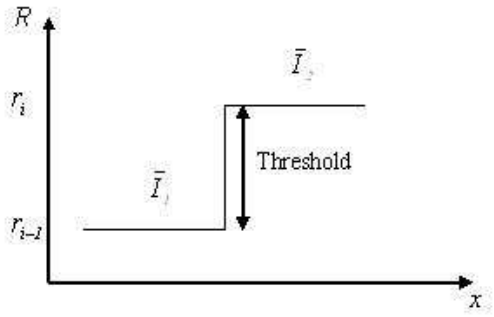


Figure 1: Unidirectional mono-edge model.

For each pixel, the ratio is calculated in different directions (horizontal, vertical, diagonals), and the minimum value of ratio is considered. A low and high ratio indicates respectively the presence of an edge and homogenous region. A pixel is affected to an edge if $r =\leq T$(T is decision threshold of detection). Else, the pixel is affected to homogenous region. Then, the performance of the operator ROA depends on the choice of decision threshold T.

The performance of the operator ROA depends on the choice of decision threshold T. It is very important to know the conditional probability distribution of ROA $P_r(r/R_1,R_2,N_1,N_2)$ to fix the false alarm rate. The probability density of the ratio for Single Look image is given by [2]

$$P_r(r/R_1,R_2,N_1,N_2) = \frac{\Gamma(N_1+N_2)\frac{1}{r}\left(\frac{N_1 r}{N_2 C}\right)^{N_1}}{\Gamma(N_1)\Gamma(N_2)\left(1+\frac{N_1 r}{N_2 C}\right)^{N_1+N_2}} \tag{3}$$

Where N_1,N_2 are the numbers of pixels in the two halves of the analyzing window. Figure 2 shows the conditional probability of ROA. In SAR image, the speckle is generally modeled as a strong multiplicative, Gamma-distributed random noise, with unity mean and variance equal to the inverse of the Equivalent Number of Independent Looks (ENIL) [3].

For a given decision threshold T and ratio contrast C, the probability of detecting an edge is:

$$P(T,C) = Prob(r > T/C) = \int_T^1 P(r/C)dr \tag{4}$$

Where $P(T,C)$ is the conditional probability of the normalized ratio.

The probability of false alarm (PFA) is the probability of detecting an edge in homogeneous region (the probability of detecting an edge in the case of the ratio contrast $C = 1$)

$$PFA(T) = P(T,1) \tag{5}$$

Relation (5) permits to compute the value of threshold T, by fixing a low PFA and the ratio of contrast $C = 1$.

The performance of the ratio edge detector depends also on the choice of the analyzing window size (size of neighborhoods). The use of the small analyzing window permits to detect the micro-edges, but in this case the effect of the speckle is not reduced. The choice of large analyzing window gives a well reduction of speckle but

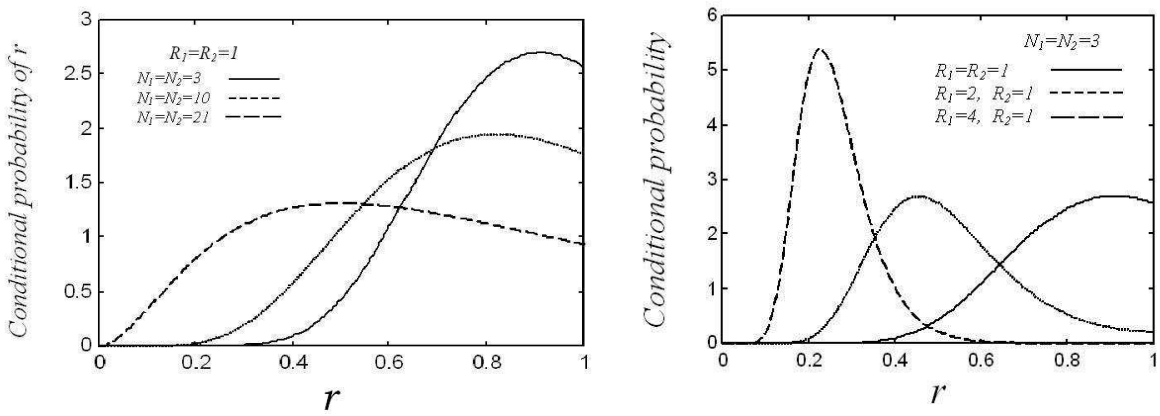


Figure 2: Conditional probability of ROA for different window sizes and reflectivity.

the micro-edges are not detected. In order to detect most edges and reduce the speckle, the ratio must be applied for different sizes of analyzing window. In our algorithm, we calculate the ratio on windows of increasing size from 3×3 to 9×9. The processes begin with computation of the ratio for the analyzing window of size 3×3. The computed ratio is compared to decision threshold and if an edge is detected we pass to the next pixel, else the ratio is computed for an analyzing window of size 5×5. We test the ratio and if an edge is not detected we pass to a window of size 7×7.The same operation is applied for each analyzing window. The process is stopped for a window size of 9×9, and if an edge is not detected for this window the considered pixel is assigned to homogenous region. To detect more edges the ROA detector is applied for all possible directions (the four usual directions are: horizontal, vertical, diagonal 1, diagonal 2). The ratio is computed for each direction and the minimum value corresponds to the direction of the most probable edge. The PFA for the ratio computed in four directions is given by [2]

$$PFA_4 = 1 - (1 - PFA_1)^3 \tag{6}$$

Where PFA_1 is the PFA of the ratio computed in one direction.

3 Ratio Of Exponentially Weighted Averages operator (ROEWA)

To reduce the influence of the speckle sufficiently, the analyzing window size must be important. A large window can contain several edges simultaneously. There is a conflict between speckle reduction and high spatial resolution. An edge detector based on multi-edge model, which is optimal in the Minimum Mean Square Error (MMSE) sense, is applied to SAR images to improve the speckle suppression and edge detection properties. Basically, it consists on the means ratio computation on opposite sides of the central pixel, with non-uniform weighting. The pixels coefficients weighting are considered as a function of the distance to the central pixel. In the case of multi-edge detector, several edges are considered in the analyzing window, Figure 3 presents the multi-edge model in one direction.

We suppose that SAR image is composed by zones of constant reflectivity. In the horizontal direction or in the vertical direction, the reflectivity R is a random process composed by segments of reflectivity, with mean μ_r and deviation σ_r. The change of reflectivity follows a Poisson distribution. That, the probability to have n jumps in the interval $\triangle x$ is given by

$$P(n/\triangle x) = \frac{(\lambda_x \triangle x)^n}{n!} e^{-\lambda_x \triangle x} \tag{7}$$

Where λ_x is the mean jump frequency.

The optimum impulse response in horizontal direction is obtained by minimizing the mean square error of reflectivity

$$f(x) = Ce^{-\alpha|x|} \tag{8}$$

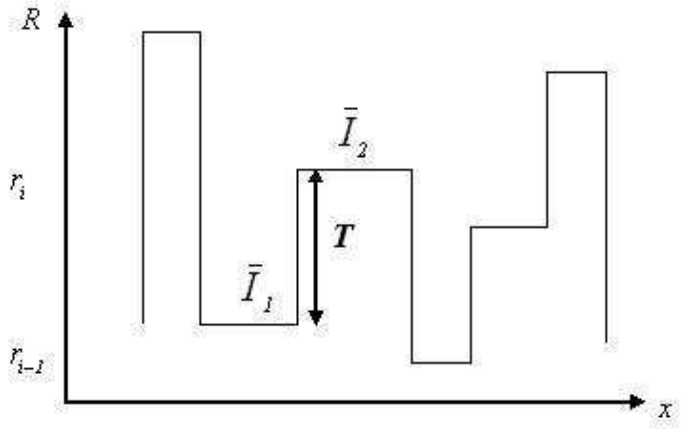


Figure 3: Unidirectional multi-edge model.

where

$$\alpha = \frac{2L\lambda_x}{1+(\mu./\sigma.)^2} + \lambda_x^2 \tag{9}$$

Figure 4 illustrates the optimum impulse response in the horizontal direction.

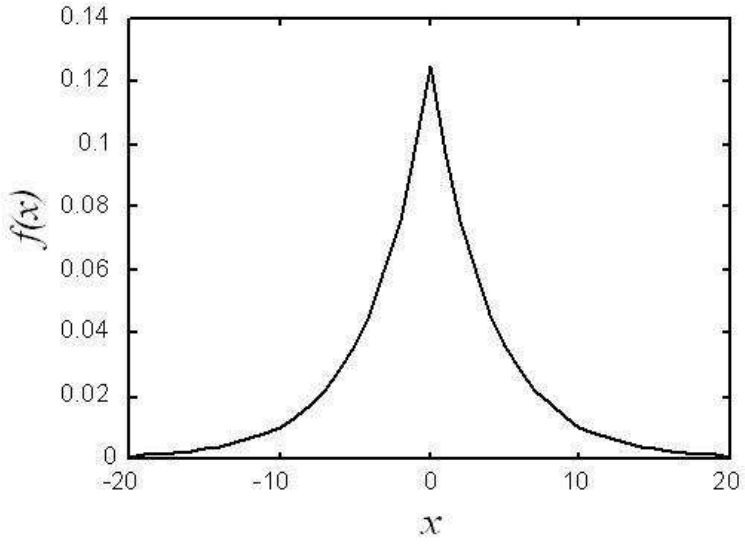


Figure 4: The impulse response of the exponential filter in horizontal direction ($\alpha = 0.25$).

In the discrete case, the filter $f(x)$ can be implemented very efficiently by pair of filters (causal and anti-causal). To detect the horizontal edge, the image is first smoothed column by column using the filter $f(x)$. Next, the means on the two opposite sides of each pixel are computed line by line independently, using the two filters causal and anti-causal. The means ratio obtained is calculated for each pixel. The ratio is normalized to lie between zero and one by taking the minimum of the ratio computed and it's inverse. The ratio in the vertical direction is computed by the same manner. The performance of this operator depends on the parameter choice α of the exponential function. The ROEWA is computed only for two directions (horizontal, vertical), the minimum of ratio indicates the presence of an edge. The ratio is compared to decision threshold, but in this case it is difficult to calculate the PFA correspond to the decision threshold.

4 Results

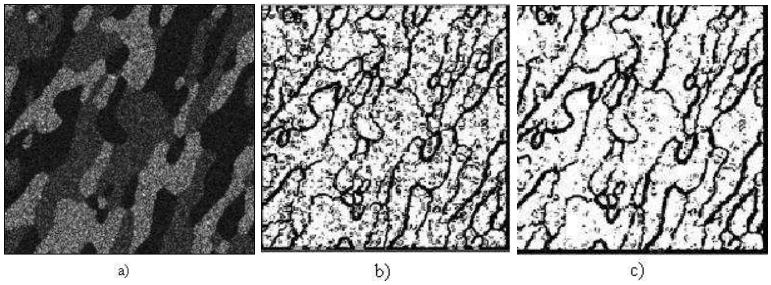


Figure 5: Edge detection of simulated image (a) speckled simulated Image (b) ROA Edge detection (c) ROEWA Edge detection

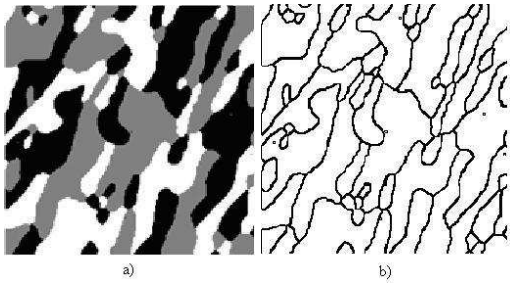


Figure 6: Edge detection of simulated image (a) simulated Image without speckle (b) ROA Edge detection

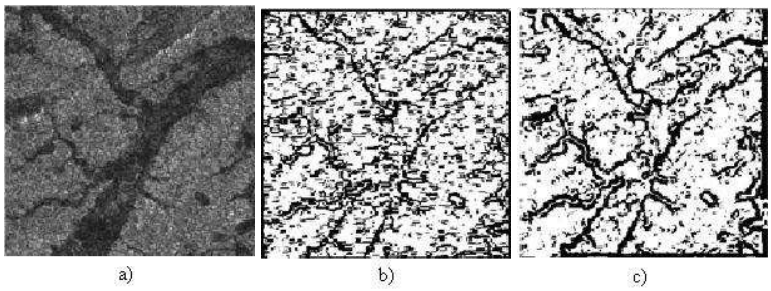


Figure 7: Edge detection of SAR image (a) ERS-1 SAR image (b) ROA Edge detection (c) ROEWA Edge detection.

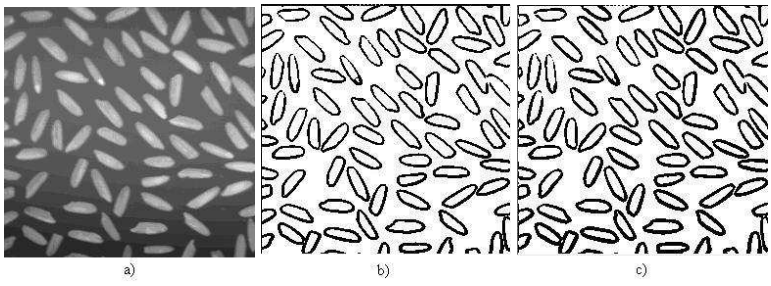


Figure 8: Edge detection of optical image (a) Optical image (b) ROA Edge detection (c) ROEWA Edge detection.

5 Conclusion

In this paper, two edge detectors are presented. These detectors are applied to simulated SAR images, and to real data of ERS-1 SAR image. The ROA operator gives best results than the usual gradient-based edge detectors. But, it is optimum only for mono-edge case. The performance of ROA depends on the choice of window size and on the decision threshold. The performance of this detector can be improved by working on different resolution. Each resolution corresponds to certain size window. The ratio is computed for different window sizes. The small window size permits to detect the micro-edges, and the large window permits a reduction of speckle. The ROEWA operator is a detector optimized for multi-edge model. The results obtained with the ROEWA detector is better than results obtained with ROA detector, when they applied to simulated SAR image and ERS-1 SAR image. The ROA and the ROEWA detectors permit to improve the edge detection in SAR images. Also, these detectors give best results when they are applied to optical images.

References

[1] A. C. Bovik, "On detecting edges in speckle imagery ". IEEE Transaction on Acoustic and Signal Processing, vol. 36, pp. 1618-1627, October 1988.

[2] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detectors for SAR images". IEEE Transaction on Geosciences and Remote Sensing, vol. 26, pp. 764-773, November 1988.

[3] A. Lopes, E. Nezry, R. Touzi, "Structure detection and statistical adaptive speckle filtering in SAR images". International Journal of Remote Sensing, vol. 14, pp. 1735-1758, 1993.

[4] R. Fjortoft, P. Marthon, A. Lopes, E. Cubero-Castan "Edge Detection in Radar Images Using Recursive Filters". IN proc. Second Asian Conference on Computer Vision [ACCV'95], Vol. 3, pages 87-91, Singapore, 5-8 December 1995.

[5] R. Fjortoft, P. Marthon, A. Lopes, E. Cubero-Castan "Multiedge detection in SAR images" in proceeding ICASSP, Volume 4. Munich, Germany, April 1997, PP. 2761-2764.

[6] J. W. Woods and Biemond "Comments on model for radar images and its application to adaptive filtering". IEEE Transaction Pattern Analysis and Machine Intelligence, Vol. 6, pp. 658-659, September 1984.

[7] J. P. Cocquerz et S. Philip, Analyse d'image : filtrage et segmentation, Paris, France : Masson, 1995.

[8] R. Fjortoft, P. Marthon, A. Lopes, E. Cubero-Castan "On optimum multiedge detector for SAR images". IEEE Transaction on Geoscience and Remote Sensing, vol. 36, pp. 793-802, May 1998.

[9] F. Galland, N. Bertaux, and P. Réfrégier "Minimum Description Length Synthetic Aperture Radar Image Segmentation". IEEE Transactions on Image Processing, Vol. 12, N° 9, pp. 995-1006, September 2003.

Mohamed Airouche, Mimoun Zelmat and Madjid Kidouche
M'hamed Bougara University of Boumerdes
Department of automatic and electrification
Avenue de l'Indépendance, FHC, UMBB, Boumerdes-35000, Algeria
E-mail: m_airou@yahoo.fr

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 150-155

Inverse Kinematics Solution of 3DOF Planar Robot using ANFIS

Srinivasan Alavandar, M.J. Nigam

Abstract: One of the most important problems in robot kinematics and control is, finding the solution of Inverse Kinematics. Traditional methods such as geometric, iterative and algebraic are inadequate if the joint structure of the manipulator is more complex. As the complexity of robot increases, obtaining the inverse kinematics is difficult and computationally expensive. In this paper, using the ability of ANFIS (Adaptive Neuro-Fuzzy Inference System) to learn from training data, it is possible to create ANFIS with limited mathematical representation of the system. Computer simulations conducted on 2DOF and 3DOF robot manipulator shows the effectiveness of the approach.
Keywords: ANFIS, manipulator, Inverse kinematics, Degree of freedom(DOF)

1 Introduction

Robot control actions are executed in the joint coordinates while robot motions are specified in the Cartesian coordinates. Conversion of the position and orientation of a robot manipulator end-effector from Cartesian space to joint space, called as inverse kinematics problem, which is of fundamental importance in calculating desired joint angles for robot manipulator design and control.

For a manipulator with n degree of freedom, at any instant of time joint variables is denoted by $\theta_i = \theta(t)$, i = 1, 2, 3,..., n and position variables $x_j = x(t)$, j = 1, 2, 3,...,m. The relations between the end-effector position $x(t)$ and joint angle $\theta(t)$ can be represented by forward kinematic equation,

$$x(t) = f(\theta(t)) \tag{1}$$

where f is a nonlinear, continuous and differentiable function. On the other hand, with the given desired end effector position, the problem of finding the values of the joint variables is inverse kinematics, which can be solved by,

$$\theta(t) = f'(x(t)) \tag{2}$$

Solution of (2) is not unique due to nonlinear, uncertain and time varying nature of the governing equations. The different techniques used for solving inverse kinematics can be classified as algebraic[1], geometric[2] and iterative[3]. The algebraic methods do not guarantee closed form solutions. In case of geometric methods, closed form solutions for the first three joints of the manipulator must exist geometrically. The iterative methods converge to only a single solution depending on the starting point and will not work near singularities. If the joints of the manipulator are more complex, the inverse kinematics solution by using these traditional methods is a time consuming.In other words, for a more generalized m-degrees of freedom manipulator, traditional methods will become prohibitive due to the high complexity of mathematical structure of the formulation. To compound the problem further, robots have to work in the real world that cannot be modeled concisely using mathematical expressions.

Utilization of Neural network (NN) and Fuzzy logic for solving the inverse kinematics is much reported[4-8]. Li-Xin Wei et al[9]., and Rasit Koker et al[10]., proposed neural network based inverse kinematics solution of a robotic manipulator. In this paper, neuro-fuzzy systems which provide fuzzy systems with automatic tuning using Neural network is used to solve the inverse kinematics problem. The paper is organized as follows, in section 2, the structure of ANFIS used is presented. Section 3 describes results and discussion. Section 4 ends with conclusion.

2 ANFIS Architecture

This section introduces the basics of ANFIS network architecture and its hybrid learning rule. Adaptive Neuro-Fuzzy Inference System is a feedforward adaptive neural network which implies a fuzzy inference system through its structure and neurons. Jang was one of the first to introduce ANFIS[11]. He reported that the ANFIS architecture can be employed to model nonlinear functions, identify nonlinear components on-line in a control system, and predict a chaotic time series. It is a hybrid neuro-fuzzy technique that brings learning capabilities of neural networks to fuzzy inference systems. The learning algorithm tunes the membership functions of a Sugeno-type Fuzzy Inference System using the training input-output data. A detailed coverage of ANFIS can be found in[11-13].

For a first order Sugeno type of rule base with two inputs x, y and one output, the structure of ANFIS is shown in Figure 1. The typical rule set can be expressed as,

Rule 1: If x_1 is A_1 AND x_2 is B_1, THEN $f_1 = p_1x + q_1y + r_1$

Rule 2: If x_1 is A_2 AND x_2 is B_2, THEN $f_2 = p_2x + q_2y + r_2$

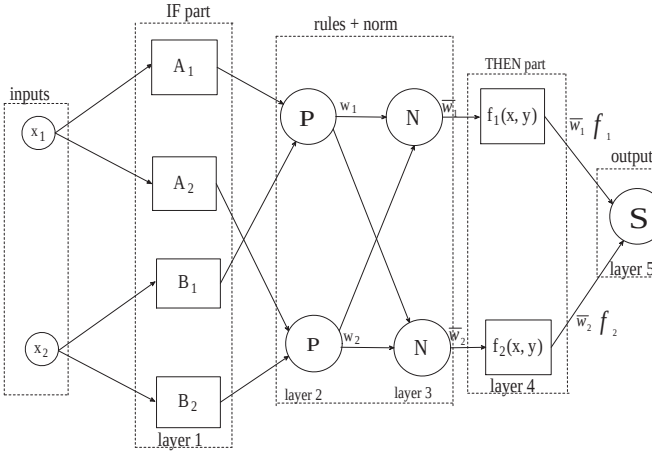


Figure 1: Structure of ANFIS

In the first layer, each node denotes the membership functions of fuzzy sets $A_i, B_i, i = 1,2$ be $\mu_{Ai}(x_1), \mu_{Bi}(x_2)$. In the second layer the T-norm operation will be done related to AND operator of fuzzy rules. Considering T-norm multiplication:

$$w_i = \mu_{Ai}(x_1).\mu_{Bi}(x_2) \tag{3}$$

In the third layer, the average is calculated based on weights taken from fuzzy rules,

$$\overline{w_i} = \frac{w_i}{w_1 + w_2} \tag{4}$$

In the fourth layer, the linear compound is obtained from the input of the system as *THEN* part of Sugeno-type fuzzy rules as,

$$\overline{w_i}.f_i = \overline{w_i}(p_ix_1 + q_ix_2 + r_i) \tag{5}$$

In the fifth layer, defuzzification process of fuzzy system (using weighted average method) is obtained by,

$$f = \sum_i \overline{w_i}.f_i = \frac{\sum_i w_i.f_i}{\sum_i w_i} \tag{6}$$

This paper considers the ANFIS structure with first order Sugeno model containing 49 rules. Gaussian membership functions with product inference rule are used at the fuzzification level. Hybrid learning algorithm that combines least square method with gradient descent method is used to adjust the parameter of membership function. The flowchart of ANFIS procedure is shown in Figure 2.

3 Simulation and Results

Figure 3(a)and 3(b) shows the two degree of freedom (DOF) and three DOF planar manipulator arm which is simulated in this work.

3.1 Two Degree of Freedom planar manipulator

For a 2 DOF planar manipulator having l_1 and l_2 as their link lengths and θ_1,θ_2 as joint angles with x, y as task coordinates the forward kinematic equations are,

$$x = l_1cos(\theta_1) + l_2cos(\theta_1 + \theta_2) \tag{7}$$

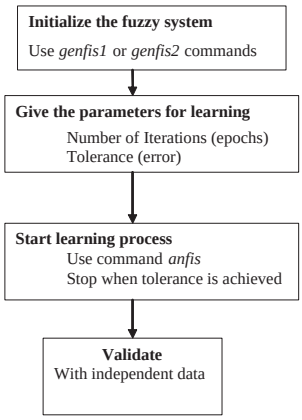
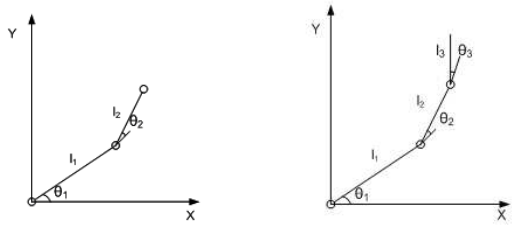


Figure 2: ANFIS procedure



(a) (b)

Figure 3: (a)Two degree of freedom (DOF) and (b)Three DOF planar manipulator

$$y = l_1 sin(\theta_1) + l_2 sin(\theta_1 + \theta_2) \tag{8}$$

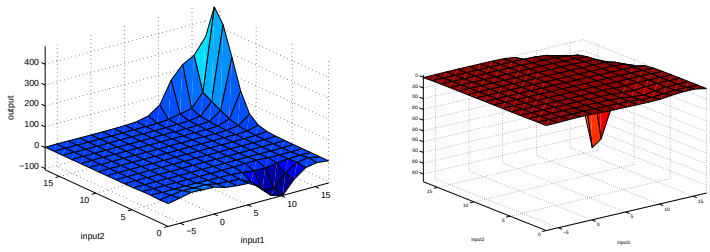
and the inverse kinematics equations are,

$$\theta_1 = atan2(y,x) - atan2(k_2,k_1) \tag{9}$$

$$\theta_2 = atan2(sin\theta_2, cos\theta_2) \tag{10}$$

where, $k_1 = l_1 + l_2 cos\theta_2$, $k_2 = l_2$,$sin\theta_2$ $cos\theta_2 = \frac{(x^2+y^2-l_1^2-l_2^2)}{2l_1l_2}$ and $sin\theta_2 = \sqrt{\pm(1-cos^2\theta_2)}$.

Considering length of first arm l_1 = 10 and length of second arm l_2 = 7 along with joint angle constraints $0 < \theta_1 < \frac{\pi}{2}, 0 < \theta_2 < \pi$, the x and y coordinates of the arm are calculated for two joints using forward kinematics. The coordinates and the angles are used as training data to train ANFIS network with Gaussian membership function with hybrid learning algorithm Figure 4 shows the training data of two ANFIS networks for two joint angles. The coordinates act as input to the ANFIS and the angles act as the output. The learning algorithm



(a) (b)

Figure 4: Training data of (a)θ_1 and (b)θ_2.

"teaches" the ANFIS to map the co-ordinates to the angles through a process called training. In the training phase,

the membership functions and the weights will be adjusted such that the required minimum error is satisfied or if the number of epochs reached. At the end of training, the trained ANFIS network would have learned the input-output map and it is tested with the deduced inverse kinematics. Figure 5 shows the difference in theta deduced analytically and the data predicted with ANFIS.

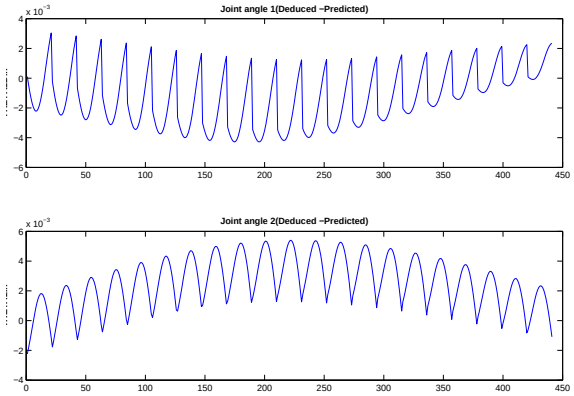


Figure 5: Difference in theta deduced and the data predicted with ANFIS trained

3.2 Three Degree of Freedom planar manipulator

For a 3 DOF planar redundant manipulator, the forward kinematic equations are,

$$x = l_1 cos(\theta_1) + l_2 cos(\theta_1 + \theta_2) + l_3 cos(\theta_1 + \theta_2 + \theta 3) \tag{11}$$

$$y = l_1 sin(\theta_1) + l_2 sin(\theta_1 + \theta_2) + l_3 sin(\theta_1 + \theta_2 + \theta 3) \tag{12}$$

$$\phi = \theta_1 + \theta_2 + \theta_3 \tag{13}$$

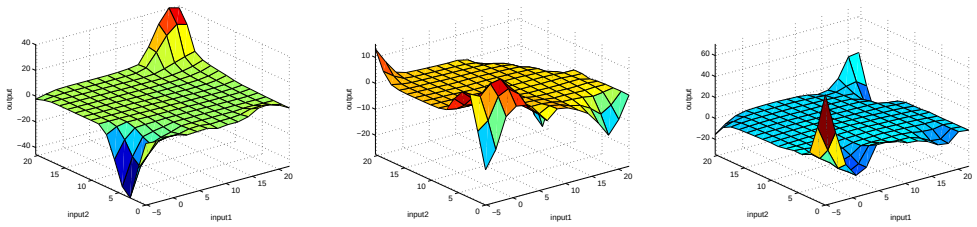
and the inverse kinematics equations are,

$$\theta_2 = atan2(sin\theta_2, cos\theta_2) \tag{14}$$

$$\theta_1 = atan2((k_1 y_n - k_2 x_n), (k_1 x_n - k_2 y_n) \tag{15}$$

$$\theta_3 = \phi - (\theta_1 + \theta_2) \tag{16}$$

where, $k_1 = l_1 + l_2 cos\theta_2$, $k_2 = l_2, sin\theta_2$ $cos\theta_2 = \frac{(x^2+y^2-l_1^2-l_2^2)}{2l_1 l_2}$, $sin\theta_2 = \sqrt{\pm(1-cos^2\theta_2)}$, $x_n = x - l_3 cos\phi$ and $y_n = y - l_3 sin\phi$. For simulation, the length for three links are l_1 = 10, l_2 = 7 and l_3 = 5 with joint angle constraints $0 < \theta_1 < \frac{\pi}{3}, 0 < \theta_2 < \frac{\pi}{2}, 0 < \theta_3 < \pi$ the same procedure is repeated. Figure 6 shows the training data of three ANFIS networks for three joint angles. Figure 7 shows the difference in theta deduced analytically and the data predicted with ANFIS.



(a) (b) (c)

Figure 6: Training data of (a)θ_1, (b)θ_2 and (c)θ_3

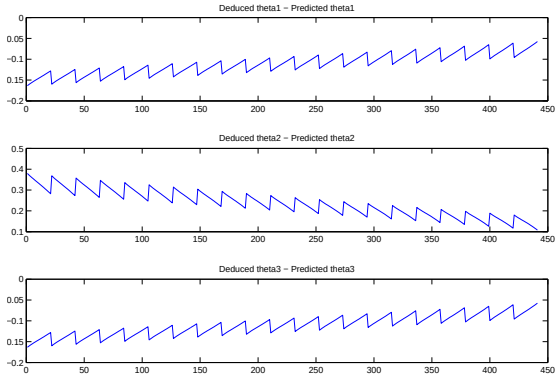


Figure 7: Difference in theta deduced and the data predicted with ANFIS trained

4 Conclusions

The difference in theta deduced and the data predicted with ANFIS trained for 2DOF and 3DOF planar manipulator clearly depicts that the proposed method results in an acceptable error. Trained ANFIS can be utilized to provide fast and acceptable solutions of the inverse kinematics thereby making ANFIS as an alternate approach to map the inverse kinematic solutions. Other techniques like input selection and alternate ways to model the problem may be explored for reducing the error further.

References

[1] J.J. Craig, *Introduction to Robotics: Mechanisms and Controls,* Addison-Wesley, Reading, MA, 1989.

[2] G.C.S. Lee, "Robot Arm Kinematics, Dynamics and Control," *Computer*, 15(12), 62-79, 1982.

[3] J.U. Korein, N.I. Balder, "Techniques for generating the goal-directed motion of articulated structures," *IEEE Computer Graphics and Applications*, 2(9), 71-81, 1982.

[4] Nedungadi, A, "Application of fuzzy logic to solve the robot inverse kinematics problem," *Proceeding of 4th World Conf. on Robotics Research*, 13, pp. 1-14, 1991.

[5] David W. Howard and Ali Zilouchian, "Application of Fuzzy Logic for the Solution of Inverse Kinematics and Hierarchical Controls of Robotic Manipulators," *Journal of Intelligent and Robotic Systems*, 23, 217-247, 1998.

[6] Sreenivas Tejomurtula, Subhash Kak, "Inverse kinematics in robotics using neural networks," *Information Sciences*, 116, 147-164, 1999

[7] Yang Ming Lu, Lu Guizhang, Li Jiangeng, "An Inverse Kinematics Solution for Manipulators," *Proceedings of IEEE*, Vol.4, 400-404, 2001.

[8] Tiberiu Vesselenyi, Simona Dzitac, Ioan Dzitac, Misu-Jan Manolescu, "Fuzzy and Neural Controllers for a Pneumatic Actuator," *International Journal of Computers, Communications and Control*, Vol. II, No. 4, pp. 375-387, 2007.

[9] Li-Xin Wei, Hong-Rui Wang, Ying Li, " A new solution for inverse kinematics of manipulator based on neural network," *Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xian*, 3(5), 1201-1203, November 2003.

[10] Rasit Koker, Cemil Oz, Tark Cakar, Huseyin Ekiz, "A study of neural network based inverse kinematics solution for a three-joint robot," *Robotics and Autonomous Systems*, 49, 227-234, 2004.

[11] J.-S. R. Jang , "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, 23(03), 665-685, May 1993.

[12] Jang, J., Sun, C., and Mizutani, E., *Neuro Fuzzy and Soft Computing,* Printice-Hall, Upper Saddle River, NJ, 1997.

[13] H. Sadjadian , H.D. Taghirad, and A. Fatehi "Neural Networks Approaches for Computing the Forward Kinematics of a Redundant Parallel Manipulator," *International Journal of Computational Intelligence* , Vol. 2, No.1, 40-47, 2005.

Srinivasan Alavandar, M.J. Nigam
Indian Institute of Technology Roorkee
Department of Electronics and Computer Engineering
Roorkee-2477667, Uttarkhand, INDIA
E-mail: seenu.phd@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 156-161

On Using Bootstrap Scenario-Generation for Multi-Period Stochastic Programming Applications

Grigore Albeanu, Manuela Ghica, Florin Popenţiu-Vlădicescu

Abstract: The bootstrap method is an extensive computational approach based on resampling and statistical estimation useful especially when only a small data set is used to predict the behavior of systems or processes. This paper proposes the usage of the bootstrap resampling for testing scenario generation methods in order to solve multi-period stochastic programming applications.
Keywords: bootstrap, multi-period stochastic programming, tree scenario-generation.

1 Introduction

Data analysis is an important topic, not only for academic area, but also for business and industrial fields [1, 2, 3, 6, 8, 13, 14, 20]. Investigating the requirements for building software to assess the accuracy of estimates not only for simple business statistical applications, but also for multi-period stochastic programming applications for long-term portfolio optimization, the bootstrap proved to be a powerful technique.

Based on some previous experience [1, 2, 3, 4] and the available scientific literature [10, 11, 12], this paper describes the usage of bootstrap methodology in scenario generation for dynamic financial analysis, and proposes two bootstrap approaches useful for solving single period and multi-period stochastic optimization problems.

The remaining part of this paper is organized as follows. Section 2, shortly, reminds the elements of bootstrap methodology to obtain the distribution of a random variable, to provide a functional relation, and to estimate the accuracy of a statistics. The third section gives a coarse overview on the dynamic financial analysis in order to identify opportunities for using the bootstrap technique. The bootstrap resampling is used during an investigation on the expected returns on portfolio. The fourth section explains the multi-stage stochastic programming and the scenario-based optimization.

Two bootstrap approaches for assessing the quality of the solution of stochastic programming problems are introduced in the fifth section together with a comparison methodology based on Hausdorff distance. The concluding part will close the presentation.

2 The basic bootstrap methodology

Let X be a random variable and F the cumulative distribution function of the variable X. The Bootstrap method, proposed by Efron [10, 11], is useful, at least, for the estimation of:

- the distribution function of a random variable $R(X,F)$;
- a functional relation $V(F)$, or
- the accuracy of a statistics s obtained from a sample $(X_1,X_2,...,X_n)$ of size n (≥ 1) from X.

For the current investigation, the accuracy, describes the variability of s when independent estimations $s(1),s(2),...,$ of the statistics s, are obtained by resampling.

The bootstrap technique uses the sample $(X_1,X_2,...,X_n)$ to obtain the sampling cumulative distribution function $F_n(x)$ in order to replace the true cumulative distribution function $F : F_n(x) = \frac{1}{n}$ cardinal $\{x_i \leq x; 1 \leq i \leq n\}$. To repeatedly simulate bootstrap samples $X^* := (X_1^*,X_2^*,...,X_n^*)$ from F_n, random number generators should be used according to the known Monte Carlo approaches [12]. Then, for each bootstrap sample, it is recalculated:

- the distribution function of the random variable $R(X^*,F_n)$;
- the functional relation $V(F_n)$ or $V(F_n^*)$; and
- the statistics $s^*()$.

The accuracy of the statistics s can be derived under an appropriate statistical inference study on the sequence $s^*()$.

The bootstrap resampling can be realised in various ways. Uniform resampling and the importance resampling are the mostly used [3, 4]. Uniform resampling assumes that the measurement (observed) values are uniformly sampled from some process, while the importance resampling algorithm assumes the generation of sampling values according to a probability distribution $\{(x_i, p_i) : i = 1,2,...,n\}$ such as every p_i is a nonnegative real number, and $p_1 + p_2 + ... + p_n = 1$. However, other approaches depending on particular applications were already proposed.

3 Empowering Dynamic Financial Analysis by Bootstraping

Dynamic Financial Analysis (DFA) is an important development based on stochastic simulation (Monte Carlo methods [12, 16, 19]) used mostly in non-life insurance and reinsurance [6, 7, 8]. However, for life insurance, the approach called Asset Liability Management (ALM) if using stochastic simulation becomes similar to DFA [15]. It is accepted now that DFA is a variant of ALM, showing a greater emphasis on both economic scenario generators and the interrelationships between assets and liabilities.

Applying this approach the insurance and reinsurance companies are able to investigate the potential impact during decisional process. Obviously, DFA borrows many concepts and methods from economics and statistics and integrating them in a powerful tool, actually implemented in software as a decision support system. DFA requires a scenario generator and a calibration procedure. After calibration, if necessary, the stochastic model can be reconsidered and improved. Also is mandatory a multivariate organization module (the logistic component), an analysis and presentation module together with a control and optimization module for improving the strategy (mainly realized in a grid environment [18]).

The scenario generator is a module which implements stochastic models for risk factors, affecting the company strategic decisions, like economic risks, liability risks, asset risks and business risks. On the first step in developing a DFA model it will be investigated the risks and the factors affecting the company results. Then, the objective functions and the projection period (the planning horizon, usually long enough) have to be chosen.

The calibration procedure is responsible for finding the suitable parameters of the models used in the scenarios generation. The difference between DFA and classical scenarios testing approach results from the usage of Monte Carlo simulation (including bootstrap) during scenarios generation and calibration.

According to [7], "the real challenge of DFA scenario generation lies in the composition of the component models into an integrated model". This task will be accomplished using both the deterministic modeling (by functional approximation) and the statistical modeling (by correlation, multivariate statistics, time series, etc). When there is only a small data set, the estimation of the model parameters (in general belonging to a high-dimensional space) asks for uncertainty diminishing. The bootstrap approach can be used for accuracy estimation as shown bellow for a simple business application. A more complex approach will be described in the next section.

In the following, an investigation on expected return on portfolio based on bootstrapping generalized regression is presented. By portfolio, according to [19] it is referred a group of financial assets. It is expected that a professional investor choose his/her portfolio so as to maximize the expected return and to minimize the risk.

Let us consider N assets, $1,2,\ldots,N$, and the measure of richness is equal to 1. The portfolio is modeled [2, 14] as the vector $x = (x_1, x_2, \ldots, x_N)^T$, where x_k represents the fraction of the unit of richness invested in the kth asset, $k = 1,2,\ldots,N$, so that $x_1 + x_2 + \cdots + x_N = 1$. The returns of the N mentioned assets are random variables $\rho = (\rho_1, \rho_2, \ldots, \rho_N)^T$, with the expected returns $r = E\rho = (r_1, r_2, \ldots, r_N)^T$, and the covariance matrix $V = (v_{ij})$, where $v_{ij} = cov(\rho_i, \rho_j), i, j = 1,2,\ldots,N$. As, usually, let us denote the diagonal elements by $\sigma_i^2 = v_{ii}$, with σ_i being the standard deviations of the returns.

For a given portfolio P, represented by weights x, the expected return on the portfolio P is $r_P = r^T x$ and the variance of the portfolio P is $\sigma_P^2 = x^T V x$, while the risk of the portfolio P is σ_P.

Usually, the return on an asset is explained in terms of a linear combination of more factors:

$$\rho_i = \sum_{j=1}^{m} \beta_{ij} F_j + \varepsilon_i, i = 1,2,\ldots,N, \tag{1}$$

where $F_j (j = 1,2,\ldots,m)$ are observed explanatory variables (like production, inflation, term structure and other economic factors), ε_i is a zero mean random disturbance (not observable), and (β_{ij}) are unknown parameters which are specific for the given asset. If there are T observations (usually gathered historical data), the regression model is $\rho_i = F^T \beta_i + \varepsilon_i$, where $\rho_i := (\rho_{i1}, \rho_{i2}, \ldots, \rho iT)^T$, $F := (Fij)$, $i = 1,2,\ldots,m, j = 1,2,\ldots,T$, is an m X T matrix,

$\beta_i := (\beta_{i1}, \beta_{i2}, \ldots, \beta_{im})^T$, and $\varepsilon_i := (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})^T$, with $E\varepsilon_{ij} = 0$ and $cov(\varepsilon_{ik}, \varepsilon_{il}) = \sigma_i^2 \delta_{kl}$, $k,l = 1,2,\ldots,T$, where $\delta_{kl} = 1$ for $k = l$ and $\delta_{kl} = 0$, otherwise.

For $T > m$ and $rank(F) = m$, the ordinary least square estimator for β_i is $\hat{\beta}_i = (FF^T)^{-1}F\rho_i$ having the covariance matrix $cov(\hat{\beta}_i) = \sigma_i^2(FF^T)^{-1}$, while an unbiased estimator for σ_i^2 is given by

$$\hat{\sigma_i^2} = \frac{1}{T-m-1}\left(\rho_i - F^T\hat{\beta}_i\right)^T\left(\rho_i - F^T\hat{\beta}_i\right). \tag{2}$$

However, when $T < m$ (small data set; short history), or $rank(F) < m$, then a least square estimator of β_i is computed using the generalized inverse [5] of F^T: $\hat{\beta}_i = (F^T)^+\rho_i$.

In this case, the confidence intervals for β_i will be obtained by bootstrapping, the generalized least square estimate being $\hat{\beta}_i^* = (F^T)^+\rho_i^*$, where $\rho_i^* = F^T\hat{\beta}_i + \varepsilon_i^*$, and ε_i^* is obtained by resampling the centered residual using the empirical distribution. For a moderate number of bootstrap steps (no more than 25), the statistics concerning the regression model will be obtained by analyzing the bootstrap results.

4 Scenario-generation methods for stochastic programming

Stochastic programming (SP) is used for modeling optimization problems dealing with uncertain parameters. Instead of single values, the parameters are described by distributions, for the single-period case, or by stochastic processes, for multi-period cases.

If the observations are collected at T different stages, in the information sets (history) $\{H_t\}_{t=1}^T$ then $H_1 \subset H_2 \cdots \subset H_T$. For $T = 2$, the single-period case is obtained. The multi-stage stochastic program with recourse at stage k uses all information provided by the collections H_t, for $t = 1,2,\ldots,k$, and will anticipate the information in H_t, for $t = k+1,\ldots,T$.

The general form of any two-stage recourse model is given by [23]:

$$\begin{array}{ll} \min & f(x) + E\left\{min_{y\in U_2}\left\{q(y,\omega)|C(\omega)x + W(\omega)y = h(\omega)\right\}\right\} \\ \text{subject to} & Ax = b, x \in U_1, \end{array} \tag{3}$$

where Ω is the support of the probability space considered for modeling, U_1 is the set of the first-stage anticipated decisions, U_2 is the set of the second-stage adaptive decisions depending on the realization of the first-stage random vector, $q(y,\omega)$ is the second-stage cost function, $\{C(\omega), W(\omega), h(\omega)|\omega \in \Omega\}$ are the random parameters with the following meaning: C is the conversion matrix between stages, W is the recourse matrix applied on second-stage adaptive decisions, and h is the second-stage information vector.

If the random vector ω has a discrete and finite distribution, with the support $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ (Ω can be seen as the scenario set), the stochastic programming problem (3) can be expressed equivalently by the following large-scale deterministic nonlinear programming problem:

$$\begin{array}{ll} \min & f(x) + \sum_{k=1}^N p_k q(y_k, \omega_k) \\ \text{subject to} & Ax = b \\ & C(\omega_k)x + W(\omega_k)y_k = h(\omega_k), \text{for all } \omega_k \in \Omega, \\ & x \in U_1, \\ & y_k \in U_2, \end{array} \tag{4}$$

where $p_k > 0$ is the probability as the kth scenario, denoted by ω_k, to be realized, and $\sum_{k=1}^N p_k = 1$.

The T-stage stochastic programming problem, an extended version of (3), consists of [15, 20, 21, 22]:

$$\min_{x_1\in U_1} f_1(x_1) + E\left[\inf_{x_2\in U_2(x_1,\phi_2)} f_2(x_2,\phi_2) + E\left[\cdots + E\left[\inf_{x_T\in U_T(x_{T-1},\phi_T)} f_T(x_T,\phi_T)\right]\right]\right], \tag{5}$$

where $\phi_2, \phi_3, \ldots, \phi_T$ are random entities, $x_t, t = 1,2,\ldots,T$ are decision variables, f_t are continuous functions, and U_t are measurable multifunctions. The entities f_1 and U_1 are deterministic, and U_1 is a non-empty set. For the linear case these entities can be written as: $f_t(x_t,\phi_t) =< c_t, x_t >$, $U_1 = \{x_1|W_1x_1 = b_1, x_1 \geq 0\}$, $U_t(x_{t-1},\phi_t) := \{x_t|C_tx_{t-1} + W_tx_t = b_t\}$, $t = 1,2,\ldots,T$, $\phi_1 = (c_1, W_1, b_1)$ is non-random (is known at the first stage), but $\phi_t = (c_t, C_t, W_t, b_t), t = 2,3,\ldots,T$ are random data.

For $T = 2$, ϕ_2 deals with random realizations. The above solution was described only for finite and discrete case.

When large values of T are necessary, the method to obtain the solution is more complex [9, 21]. Even, for $T = 2$, if ϕ_2 has an infinite (or very large) number) of possible realizations the basic approach is to represent this distribution by scenarios.

The following questions arise: a) how many scenarios are necessary and how to create these scenarios; b) how to solve the very large-scale programming problem equivalent to (4) or more difficult required by (5) c) How to assess the quality of the obtained solution when refer to the true objective value and the true optimal solution.

The most used approach to address the first two of the above requirements deals with the sample average approximation (SAA) method [22], where scenarios are created by the Monte-Carlo method. A three-period tree describes a stochastic process having two realizations in the first period, and three realizations per vertex in the last two periods. These scheme can be easy extended to "T-period" scenario trees. However, if the probability distribution is not fully known, or we have some theoretical information about the distribution family, or only we have data, a bootstrap approach can be used to assess the quality of the obtained solution. The next section will provide a short description of the bootstrap approach for assessing optimal solutions of stochastic programming problems (SPPs))based on multi-period scenarios trees.

5 Bootstrap approach for multi-period scenario trees

Two variants can be proposed for bootstrapping. We called these approaches uni-tree and multi-tree bootstrapping.

The first of them (uni-tree) starts with a scenario tree, and compute an initial solution s_0. Then by using uniform branch resampling (for $i = 1,2,\ldots,B$), a new SPP' solution s_i is obtained (by a PC (Personal Computers)) clustering optimization method [18]). A statistical analysis can be used to study the variability of the solution based on the chain $s_0, s_1, \ldots, s_B$. When a theoretical solution s^* is available, a comparison and statistical inference can be applied. When no previous solution is available, computational statistics can be used to obtain an estimation for the SPP' solution.

The second one create $B+1$ scenario-trees $TR_i, i = 0,1,\ldots,B$ by resampling historical data coming from the sets $H_1, H_2, \ldots, H_T$. For every tree i, a SPP' solution, denoted by STR_i, is obtained (using the PC clustering). Then a statistical analysis can be used to address the quality of the considered SPP.

Let d be any distance based on a norm $||.||$ ($||.||_2$, $||.||_1$, $||.||_\infty$, etc.), where $||x||$ is a norm defined on the space containing x. Let $S_1 = (s_0, s_1, \ldots, s_B)$ and $S_2 = (STR_0, STR_1, \ldots, STR_B)$. A comparison between S_1 and S_2 is better to be realized based on the Hausdorff distance:

$$d(S_1,S_2) = \sup_{i=0,1,\ldots,B} \left\{ \inf_{j=0,1,\ldots,B} \left\{ d(s_i, STR_j) \right\} \right\}. \tag{6}$$

Applying the two methods for practical SPP conclude with a small difference between the two chains of solutions. Comparing against the known solution, the multi-tree approach proves to be better. This proves to be quite equivalent with SPP solving based on a scenario-tree with an increasing number of descendent vertex.

The proposed method is an alternate approach to the procedure described by [17], and, even computational intensive, the method provides a solution with a large confidence (quite close to the feeling of the application behavior).

6 Summary and Conclusions

The bootstrap approach was used by authors for different fields of application. This paper extends the current approaches for solving stochastic programming problems by the assessment of the quality of the SP' solution by bootstrap methodology. Of course, the complexity of the new proposal depends on the complexity of the SP's problem in discussion. A PC-cluster solution is required (at least) and no more than ten bootstrap scenario trees are required for a high quality assessment.

Acknowledgment. The first two authors acknowledge the support of the Research Center in Mathematics and Computer Science from Spiru Haret University, while the third author investigates stochastic optimization problems in the framework of the Internal Research Project of the UNESCO Chair in Information Technologies at University of Oradea.

References

[1] G. Albeanu, H. Madsen, M. Ghica, P. Thyregod, F. Popenţiu-Vlădicescu, "On using bootstrap methods for understanding empirical loss data and dynamic financial analysis," *The Seventh Annual Conferece of The European Network of Business and Industrial Statistics*, September 24-26, 2007, Dortmund, Germany, CD-ROM.

[2] G. Albeanu, M. Ghica, "Bootstrap simulation models for probabilistic approaches in environmental economics," *The 6th Workshop on Mathematical Modelling of Environmental and Life Sciences Problems*, September 5-9, 2007, Constanţa, Romania (in press).

[3] G. Albeanu, F. Popenţiu, "On the Bootstrap Method: Software Reliability Assessment and Simultaneous Confidence Bands," *Annals of Oradea University, Energetic Series*, Vol 7, No. 1, pp. 109-113, 2001.

[4] G. Albeanu, "Resampling Simultaneous Confidence Bands for Nonlinear Explicit Regression Models," *Mathematical Reports*, Vol. 50, No. 5-6, pp. 289-295, 1998.

[5] A. Ben-Israel, T. N. E. Greville, Generalized Inverses: Theory and Applications, *Wiley*, New York, 1977.

[6] A. Bergbauer, V. Charez, T. Fischer, R. Perera, A. Roehrl, S. Schmiedl, "Back to the future: Dynamic financial analysis (DFA) for decision making", 2004, http://www.approximity.com/papers/dfa_wp4en.pdf.

[7] P. Blum, M. Dacorogna, "DFA-Dynamic Financial Analysis," *Encyclopaedia of Actuarial Science*, John Wiley & Sons, 2004.

[8] R. A. Derrig, K. M. Ostaszewski, G. A. Rempala, "Applications of resampling methods in actuarial practice," *Proceedings of Casualty Actuarial Society*, Vol. 87, pp. 222-264, 2000.

[9] M. Dyer, L. Stougie, "Computational complexity of stochastic programming problems," *SPOR-Report 2005-11*, Dept. of Mathematics and Computer Science, Eindhoven Technical University, Eindhoven, 2005.

[10] B. Efron, "Computer-Intensive Methods in Statistical Regression," *SIAM Review*, Vol. 30, No. 3, pp. 421-449, 1988.

[11] B. Efron, R. J. Tibshirani, An introduction to the bootstrap, *Chapman and Hall, New York,* 1993.

[12] G. Fishman, Monte Carlo: Concepts, Algorithms, and Applications, *Springer, Berlin,* 1996.

[13] M. Ghica, A risk exchange model with a mixture exponential utility function, *Annals of Bucharest University*, Mathematics and Informatics Series, Vol. LV, No. 1, 2006.

[14] M. Ghica, Optimal portfolio in finance and actuarial science, *Annals of Spiru Haret University*, Mathematics and Informatics Series (in press).

[15] N. Hibiki, "Multi-period Stochastic Optimization Models for Dynamic Asset Allocation," *Journal of Banking and Finance*, Vol. 30, No.2, pp. 365-390, 2006.

[16] P. Jǎckel, Monte Carlo methods in finance, *John Wiley & Sons*, Chichester, 2002.

[17] M. Kaut, S. W. Wallace, "Evaluation of scenario-generation methods for stochastic programming," *SPEPS*, Working Paper 14 (http://edoc.hu-berlin.de/series/speps/2003-14/PDF/14.pdf), 2003.

[18] J. Linderoth, S. J. Wright, "Computational Grids for Stochastic Programming," *Optimization Technical Report 01-01*, Computer Sciences Department, University of Wisconsin-Madison, 2002 (http://www.cs.wisc.edu/ swright/papers/uwopt-0101.pdf)

[19] A. G. Malliaris, W. A. Brock, Stochastic methods in economics and finance, *Elsevier Science B.V.*,Amsterdam, 1982.

[20] J. M. Mulvey, "Multi-stage Optimization for Long-term Investors," *Quantitative Analysis in Financial Markets*, Vol. 3 (ed. M. Avellaneda), World Scientific Publishing Co., Singapore, pp. 66-85, 2001.

[21] A. Shapiro, "On Complexity of Multistage Stochastic Programs," *Optimization Online*, 2005: http://www.optimization-online.org/index.html.

[22] C. Swamy, D. B. Shmoy, "Sampling-based Approximation Algorithms for Multi-stage Stochastic Optimization," *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pp. 357-366, 2005.

[23] L.-Y. Yu, X.-D. Ji, S.-Y. Wang, "Stochastic Programming Models in Financial Optimization: A Survey," *AMO-Advanced Modeling and Optimization - An Electronic International Journal,* Vol. 5, No. 1, pp. 1-26 (http://www.ici.ro/camo/journal/vol5/v5a1.pdf), 2003.

Grigore Albeanu
Spiru Haret University
Department of Mathematics and Informatics
13, Ion Ghica Str., 030045, Bucharest, Romania
E-mail: galbeanu@gmail.com

Manuela Ghica
Spiru Haret University
Department of Mathematics and Informatics
13, Ion Ghica Str., 030045, Bucharest, Romania
E-mail: nimanuela@yahoo.com

Florin Popenţiu-Vlădicescu
University of Oradea
UNESCO Chair in Information Technologies
1, Universităţii Str., 410087, Oradea, Romania
E-mail: popentiu@imm.dtu.dk

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 162-167

Semantic Integrity Control in the Database Layer of an e-Health System

Lenuţa Alboaie, Diana Gorea, Victor Felea

Abstract: The paper presents a secure and efficient e-Health platform using the actual paradigms and standards like SOA and Web services. We focus on the semantic integrity aspect in the database layer of the proposed system. We also provided an overview of the situations that can affect consistency, as well as the the way we approach each of these.
Keywords: semantic, integrity control, e-Health, design, platform, SOA, Web services, OpenID, decision support

1 Introduction

At this moment, many of the existing e-health system are difficult to use in emergency situations by its users: patients, medical staff, and auxiliary personnel. The main reason is that the architecture of those systems consists of several Web applications residing on different sites, often using certain incompatible authentication procedures and access policies. The patients and medical staff must perform multiple logins and additional tasks, and they must spend time to search (vital) information located in many places. Furthermore there are only a few existing systems that meet the scalability requirements.

In order to develop a complex and useful e-health system we need to employ paradigms and/or standards such as SOA (Service Oriented Architecture) and Web services.

This paper describes focuses on the database layer of our e-Health system, called Telemon [16] , that conforms to the above mentioned architecture. It continues previous work described in [2], in which the general architecture of Telemon is presented. The current work focuses on the semantic integrity control [1] issue in the database layer of the system. This aspect is treated by exploiting the services that Telemon's SOA architecture offers.

2 Functional and architectural perspective of Telemon e-health system

In this section we describe the functionality and the general architecture of Telemon e-Health system.

2.1 Telemon - expected outcome

Telemon's goal is to offer services to both the patients and the doctors. The system will offer to patients the possibility to access emergency services and to contact medical staff directly (e.g., family doctor or nurse). Also, from hospitals, policlinics and ambulances, the medical staff (doctors, nurses) can permanently supervise the patients' status (especially those with special health problems such as chronic diseases - e.g. hypertension).

The patient's record file will be supplied dynamically via certain proper devices (in some cases, they will be wireless such as mobile phones, PDAs etc.).

The system provides a set of Web services for the patients and the medical staff. We list below some of these services: creating users profiles, managing users profiles, accessing information about patients (for example, disease history), accessing information about similar cases; processing (bio)signals, producing alarms regarding the patient's physical status, searching the nearest medical unit in emergency situations, information related to users (e.g., patients, doctors, medical personal, system administrators). All these data is stored in a distributed manner by using database services.

2.2 Telemon Architecture

Overview

Telemon is intended to allow real time patient monitoring, transferring results to local sub-systems, which at their turn will update the central system. This update will be done depending on some factors that we will discuss in next section.

The general architecture of our health system platform is depicted in Figure 1.

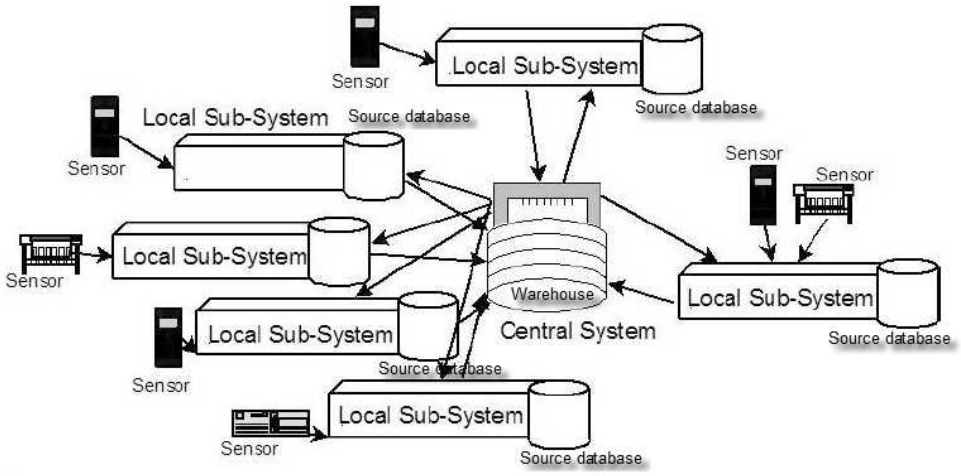


Figure 1: Telemon architecture

The sensors are devices that acquire information (e.g. SpO2 - oxygen saturation, EKG) from patients. They have the ability to transmit information to the proximal local sub-systems. In the same time, the sensors information can be accessed from various mobile devices such as: ambulance terminals, PDAs, laptops etc.

Medical staff can access information from the local sub-system based on an authentication service, as well as complete information in the central system.

In Telemon the information can be easily accessed using an automated authentication and authorization system that allows a user that is authenticated and authorized on a local sub-system, to be automatically authenticated and authorized on the central system. The procedure works the other way around as well, that is, the user that is authenticated in the central system will be automatically authenticated in all the local sub-systems. The authorization to access information is based on user types and user groups (family doctors, specialist doctors, statisticians, researchers, patients). Furthermore, logging out from a local sub-system involves logging out from the central system.

To achieve this we use a single sign-on (SSO) mechanism - which according to [3] is a system whereby a single action of user authentication and authorization can permit a user access to all computers and systems where he has access permission to, without the need to enter multiple passwords. Single sign-on reduces human error, a major component of systems failure and is therefore highly desirable for this kind of e-Health systems. In the future work we shall propose an OpenID module (which is a decentralized single sign-on system) [4] that will ensure the whole authentication process used in Telemon.

Ensuring a right authentication and authorization module contributes significantly to the maintenace of the database integrity.

Architectural view on the components

We outline the general structure of the components - local subsystems and the central system, both of which conforming to SOA principles - and for each of them we decompose the architecture in several levels.

Each of the components consists of the following layers:

User-interaction Layer - from a user interaction perspective, our system is enriched with accessibility capabilities. The provided user-interface supports users having various disabilities, according to WAI (Web Accessibility Initiative) [15].

The performed activities must be effective, efficient and secure, both at the conventional Web browser level and at the mobile application level.

An important feature will be the support offered by the GIS [5] web services, which will offer to authenticated patients information about the known pharmacies, clinics, medical offices in their proximity.

Telemon core level - from the technical point of view the system conforms to the SOA paradigm. The term Service Oriented Architecture [6], [7] refers to the design of a distributed system. SOA is not a new technology. It is a novel design methodology and architecture aimed at maximizing the reuse of multiple services (possibly implemented on different platforms and using multiple programming languages).

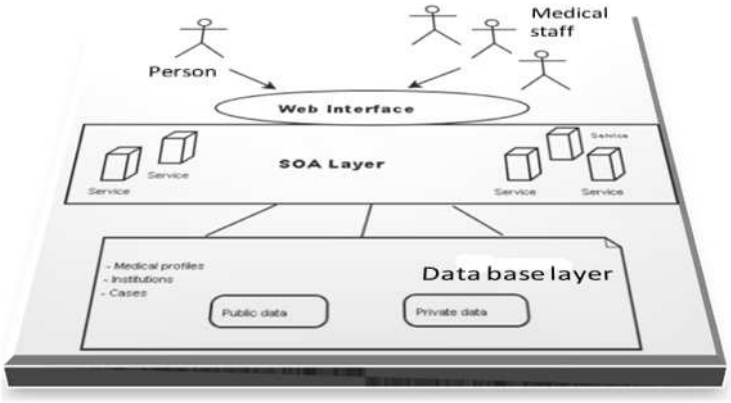


Figure 2: A sliced overview of a component

Telemon core level consists of three main modules:

- Service module: its function is matching up the requester to the provider (e.g. a doctor monitors the patient condition and wants to access his medical record. For that, he accesses a service called GetPatientRecord. The request is delivered to the Middleware, where the service module matches it to the corresponding service provider) ;
- Workflow module: its function is to do the choreography. In other words, this module does the coordination task;
- Registry module: its role is to easily identify the most appropriate service among all the existing services that provide the same function. Therefore we must have a strict evidence of the available services. The registry module can maintain information like:

  Interface descriptions (e.g., WSDL) [13], meta-data that can represent relations between services, service level agreements.

The aimed architectural solution is a multi-platform one, loosely coupled, facilitating the integration of applications, services and systems at the Web level.

The Database level. Telemon comprises two types of databases. The one located at the sub-system level is an operational data base that records data sent by the sensors (they are named *source databases*). The other is a *warehouse* and is located at the central system level.

Data warehousing refers to a set of technologies for improving the decision making processes. In this respect, Telemon includes a Decision Support Module (DSM), which will use a system named DeVisa, described in [10]. The system stores and processes prediction models using XML technologies [9] and is intended to provide feasible solutions for dynamic knowledge and decision support integration into applications. A use case would be when a doctor needs assistance in taking a decision about a treatment for an emergency situation. DeVisa provides prediction services that will be integrated as a module in Telemon, due to the chosen SOA architecture.

The core functionality of DeVisa is available via web services. The prediction models stored in DeVisa are built on top of the central warehouse, and uploaded in DeVisa via web services.

3 Database integrity

In this section we describe the semantic integrity control in Telemon. Semantic integrity [1] control(SIC) is a technique meant to ensure database consistency. Other techniques that assures database consistency are: concurrency control, reliability and protection.

We can maintain the database consistency by enforcing a set of constraints. There are two kinds of constraints:

- structural constraints (e.g., unique key constraint in the relational model)

- behavioral constraints

The structural constraints can be expressed using compiled individual assertions of relational calculus, as depicted in [1]. The behavioral constraints typically involve data from multiple sites (source and/or central databases) and are expressed using predicates as described in the following sections.

There are two main cases to consider when tackling the consistency issues in Telemon: the warehouse consistency and the source database consistency. In general the consistency in databases may be affected after one ore more updates take place. In section 3.1 we describe the possible types of updates that may happen in Telemon and their consequences. In section 3.2 and 3.3 we will focus on a set of assertions representing mainly behavioral constraints that contribute to enforcing Telemon's database layer integrity.

3.1 Update scenarios

This section presents the context and the scenarios that can affect the consistency of the database. As we stated in 2.2 the database level of Telemon is formed by two types of databases: source databases and central system warehouse.

The updates for the source databases originate from two sources:

- continuous data sent by the sensors
- the data that are operated into the system by the medical staff

The source databases feed the warehouse with data. Our system, partially decentralized, ensures the data transfer from the local sub-systems to the central system.

To optimize the data traffic we designed two different types of warehouse updates: instant updates and periodical updates. Periodical updates are done on a regular basis and they apply to all the data. The instant updates depend on an emergency threshold, which refers to the severity of the patient's condition. In the cases where the severity is above the emergency threshold, an instant update is applied in the warehouse. This is crucial when an emergency case occurs and the doctor can consult the specialists that are logged in the central system at that time. The specialists logged into the central system are automatically logged on to the local sub-system (using the openID module). In this way they can interfere in the patient treatment directly in the local sub-system in whose proximity the patient is situated. The information concerning the treatment syncs back with the central system via the periodical or instant updates. Furthermore, when there is a serious case of a similar type in another sub-system, the doctor who is logged on there consults the warehouse to get the previous treatment in order to take a decision.

3.2 Source database consistency

This section discusses the situation in which the consistecy of the source database can be affected and we establish the assertions to enforce integrity.

In Telemon the access to a patient record can be granted to more than one doctor. The interesting situation is when two or more doctors write observations or treatments in the patient record in the same time. In this case we resolve the concurrency control through the use of a priority mechanism.

In this sense, we established a priority access (for the write action) to the patient records. The doctor that is conducting the consult with a specific patient has the highest priority, so he writes directly in the patient's record, after the record is locked. Should another doctor need to write something in the patient's record in the same time, he actually writes in a temporary buffer. After he writes down all the observations and the patient's record is not locked anymore, the information is dynamically added. In addition to the priority mechanism described above, we consider that updates to the patient record are done conforming to a time-stamp, which represents the moment in time when each of the doctors accessed the patient record.

In the above scenario we stated that the mechanism that formalizes the access priority to the patient's record is based on behavioral constraints. In addition there are structural constraints associated with the relational database model used by the local sub-systems.

3.3 Warehouse consistency

The source database semantic integrity is solved by means of concurrency control mechanisms. In the case of the warehouse there are a few possible scenarios that can affect the consistency and for each of them we will identify the assertions that form the behavioral constraints.

In the regular case of periodical updates it is not necessary to specify Telemon behavioral constraints, because the updates take place according to the general integrity control principles in a database.

Therefore we focus only on the case in which the warehouse is updated from two or more source databases. One possible use case is when a patient migrates in another area than his/her usual sub-system area. Let's assume that patient A is registered in the proximity of the sub-system S1 and the patient travels to the proximity of another sub-system: S2. Thus the S2 sensor will send the corresponding data to the S2 sub-system from the moment it has detected the patient. Furthermore, S2 will trigger a workflow that comprises: a dynamic service from Telemon core level (see [2]) which has the ability to find the base subsystem of the patient - which is S1 in this case and notifies S1 of the fact that the patient is in S2's proximity.

In this circumstance we distinguish two cases:

- During his/her stay in the proximity of S2 the patient does not need emergency medical assistance. Nevertheless Telemon is in continuous surveillance of the patient. After storing the information received from the sensor S2 invokes an update service that has as input the reference to S1 (obtained by the previous service) and transfers the patient's data to S1. Hereby we identify a behavioral constraint that states that while A is in S2's area, S2 stores his records and afterwards it enforces the update between the S2 and S1. This constraint avoids unnecessary updates of the central system, because the periodical updates (see 3.1) are done by S1 after it receives the data.

- During his/her stay in the proximity of S2 the patient needs medical assistance. When S2 perceives emergency information from the sensor it triggers an emergency update procedure that updates the patient data in his/her base sub-system (S1). The doctor that assists the patient logs in S2 and he is automatically logged in S1 (OpenID) and can interact with the A's personal doctor from S1 in order to prescribe the most appropriate treatment given the circumstances. In this case a set of assertions that state that if S2 detects an emergency data it triggers an emergency update to S1.

The assertions form the base of the database consistency control and in conjunction with the authentication and authorization module we achieve the Telemon semantic integrity control.

4 Conclusions

This paper presented the semantic integrity aspect in an e-Health system that conforms to Service Oriented Architecture. The system contains component such as: sensors, local sub-systems and central system. The sensors acquire information from patients and send it to the proximal local sub-systems, which update the central system. The central system stores the data in a warehouse and provides services such as: statistics, decision support, search etc. The security aspects are solved by using a OpenID mechanism, which will be detailed in our future work. We provided an overview of the situations that can affect Telemon's database consistency and the way we approached each of the cases.

References

[1] M. Tamer, P. Valduriez, *Distributed Database Systems*, Prentice Hall, 1999.

[2] L. Alboaie, S. Buraga, "Service-oriented architecture for health systems," *The Proceedings of the national symposium with international participation e-Health and bioengineering- EHB2007,* Nr.2, 2007

[3] *Single Sign-On* , http://www.opengroup.org/security/sso/

[4] *OpenID*, http://openid.net/foundation/

[5] *Guide to Geographic Information Systems* , http://www.gis.com/index.html

[6] L. Alboaie, S. Buraga, *Web Services*, Polirom, 2006

[7] T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design,* Prentice Hall, 2005

[8] S. Buraga, *XML Technologies*, Polirom, 2006

[9] T. Bray et al. (eds.), *Extensible Markup Language (XML) 1.0 (Fourth Edition),* W3C Recommendation, 2006: http://www.w3.org/TR/XML/

[10] Gorea, D. (2007), "Towards storing and interchanging data mining models," *Proceedings of the 3rd Balkan Conference in Informatics,* volume 2, pages 229Ð-236.

[11] *ARTEMIS,* http://www.srdc.metu.edu.tr/webpage/projects/artemis/publications.html

[12] *semanticHealth,* http://www.semantichealth.org/

[13] *WSDL - Web Services Description Language*, http://www.w3.org/TR/wsdl

[14] B. Daum, U. Merten, *System Architecture with XML*, Elsevier Science, 2003

[15] * * *, *World Wide Consortium*, http://www.w3.org/

[16] * * *, *Telemon Project*, http://thor.info.uaic.ro/ ˜telemonfcs/

Lenuţa Alboaie, Diana Gorea, Victor Felea
University Al. I. Cuza Iaşi
Faculty of Computer Science
St. G-ral Berthelot nr 16
E-mail: {adria,dgorea,felea}@infoiasi.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 168-172

A Development Process for Enterprise Information Systems Based on Automatic Generation of the Components

Adrian Alexandrescu

Abstract: This paper contains some ideas concerning the Enterprise Information Systems (EIS) development. It combines known elements from the software engineering domain, with original elements, which the author has conceived and experimented. The author has followed two major objectives: to use a simple description for the concepts of an EIS, and to achieve a rapid and reliable EIS development process with minimal cost. The first goal was achieved defining some models, which describes the conceptual elements of the EIS domain: entities, events, actions, states and attribute-domain. The second goal is based on a predefined architectural model for the EIS, on predefined analyze and design models for the elements of the domain and finally on the automatic generation of the system components. The proposed methods do not depend on a special programming language or a data base management system. They are general and may be applied to any combination of such technologies.
Keywords: Enterprise Information System, Conceptual modeling, Software development process, Automatic Generation of Components

1 Introduction

Enterprise Information Systems (EIS) are complex systems containing many components, which interact. The components correspond to the main concepts we may identify in the business domain of the enterprise. There are a lot of similarities in the developing process of the components. If we identify the categories of concepts and create models for them and for their interactions, we'll face with a lot of repeatable work we must do for developing these components. The automatic generation of the components is a very useful and attractive issue that we may apply in the case of the EIS development.

This article tried to identify the steps we must follow for achieving the goal of generating the components of an EIS. First we must identify the categories of concepts of the business domain. We must then develop patterns for each category. A language for specifying the concepts is also very useful. For each concept we must do such a specification. Starting from each pattern type and the concept specification, we must then create a component type generator. Now we can execute the component generator for each concept of the domain. In addition, the developer disposes of a predefined system architecture for the EIS and some predefined components. The developer can also add its own features to the system.

The article proposes the categories of concepts for an EIS, and gives an example of a pattern for a concept category.

2 The Modeling Elements for the EIS Concepts

We'll consider an EIS as an *event* processing system. The events appear, in the enterprise activity at different moments, and they are in general, descriptions of the usage and/or transformation of the enterprise resources, at a given moment. Examples of events may be the receiving of an order from a customer, for products or services delivering; the employment of a new person in the enterprise; the emitting of an invoice by a supplier, for its services done in a certain period, etc.

We'll associate to an event a synthetic description which contains identification elements, the initiator of the event etc., and also one or more detailed descriptions of the event, defined by sequences of *actions*. The events, both by synthetic description and by detailed description, refer to a set of concepts from the enterprise activity domain, which we'll name *entities*. An entity may be: a customer, a product, a supplier, an employee, etc. The entities can be things, beings or abstract notions, involved in the events. The enterprise resources are also entities.

Between the concepts presented above may exist association relations.

The *state of an entity* is a quantitative and qualitative evaluation of an entity at a given moment. The stock of a product at a given moment is the state of a product type entity at that moment, containing an evaluation of its quantity and value. Usually, we are interested about the state of the enterprise resources at a given moment. The entities, the events and the states have some characteristics which we'll name *attributes*. An attribute may

take values from a predefined set, which we'll name *attribute-domain*. The attributes can also represent references within associations. For example, the customer entity has a name, an unique identification code, an address, etc; the order emitting event has an identification number, an emitting date, an identification customer code, etc.; the action associated to an emitting order event may have as attribute the product code, the measure unit, the quantity, etc.

By analogy between an event description and a natural language compound sentence, we may consider an action the equivalent of each sentence from the compound sentence, the nouns may correspond to entities and determinants may be similar to the entity and action attributes.

We define the state of the system at a moment as the set of the system entities states at that moment.

The instances of the concepts will be grouped in sets. So, we'll have event-sets, action-sets, entity-sets and state-sets. For example, the trading products will form an entity-set, which we'll name Products, the employees of an enterprise form the entity-set Employees, etc. Similarly, the orders emitted by the customers we'll group into an event-set named Orders. Also, the orders content from the customers we'll group into an action-set called Orders-Content. Finally, the products stock, considered at different moments (for example at the beginning of every month) will form a state-set which we may name Stocks.

By *system conceptual elements* or shortly *system-elements* we'll understand: the entity-sets, the event-sets, the action-sets, the state-sets and the attribute-domains. The system-elements will play a very important role in the EIS developing process.

3 Predefined Analyze Models, for the System-Elements

Once identified, a system-element dispose of a predefined model for the analyze and design phases. The user will perceive a system-element by means of a predefined user-interface associated with that system-element, this meaning a predefined structure and a predefined use-case. The parameter parts of the predefined models are specified using a system-element description language (SEDL).

For example, an entity-set has the user-interface structure (IES) presented in figure 1. The user-interface is an object, which contains data and operations, available to the user, if it has certain rights.

The entities have the *presentation attributes* A_1, A_2, ..., A_p, for the user. The user-interface contains the predefined operations O_1, O_2, ..., O_m. As operation examples we may have: the detailed view for the current entity, entering a new entity, edit the attributes values of an entity, etc.

The I_{ES} object contains four parts: the tabular image, the detailed image, the filtering elements and the operations.

The tabular image contains the entity-set. The columns correspond to the presentation attributes, and the rows, to the entities. The tabular image has an *entity-selector*, which indicates the *current-entity*.

The detailed image contains the presentation attributes values for the current entity. The detailed image is used when demanding operations such as: detailed viewing, enter a new entity and modifying an entity.

The filtering elements enable the view of a subset of the entity-set. They correspond to some presentation attributes (filtering attributes) of the entity-set and may be shown as combo-boxes (L_i) or list-boxes (L). By user selection of some values from these lists, the tabular image will be filtered to those entities, which have the filtering attribute values equal to the selected values.

The use-case of such an user-interface may be described by a state transition diagram, as in the figure 2. There is synthetically described, the interaction between the user and the user-interface object I_{ES}, for an entity-set.

The events: new entity, delete the current entity, edit the current entity, save and cancel, are triggered by clicking the corresponding operation buttons of the user-interface. The diagram contains only the main elements of the interaction between the user and the interface I_{ES}. There are missing from the diagram, the detailed viewing demand, the configuration of the updating user rights, tabular image configuration commands (change columns order, resize column, hide columns etc) sorting tabular image, finding entities in the tabular image, more filtering facilities for the tabular image etc.

Similarly, there are defined models for the other system-elements: attribute-domain, event-set, action-set and state-set.

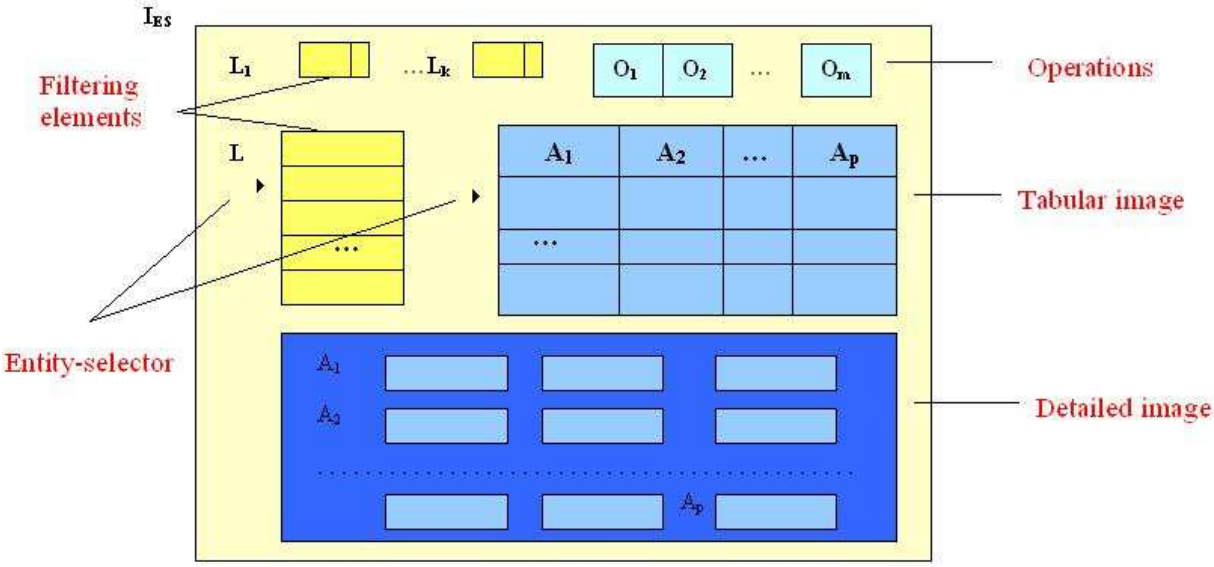


Figure 1: The structure of the entity-set user-interface, I_{ES}

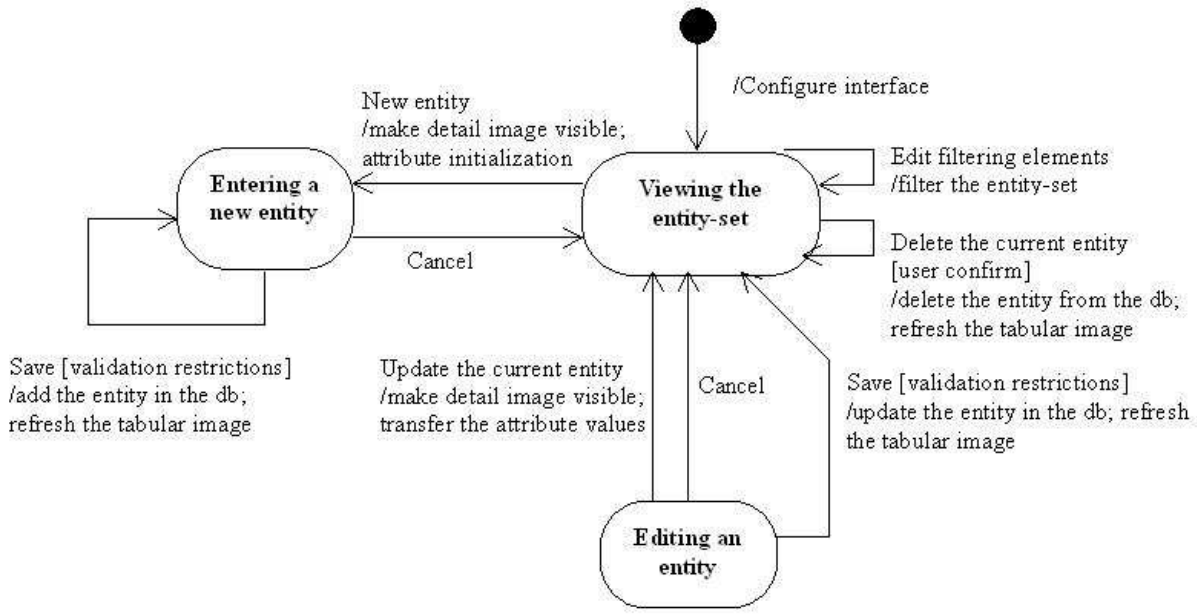


Figure 2: The use case description for the I_{ES}

4 The Design of the System-elements

The design of a system-element consists in specifying its attributes and the properties of the attributes, using a system-element description language (SEDL). In general, each system-element, need one relation in a relational data base.

We'll give again as a system-element example, an entity-set, noted ES. For the persistence purpose, we'll associate a relation, R_{ES} from the data base, with the schema R_{ES} (A_1, A_2,..., A_n). We'll consider that the relation schema is in the third normal form. We'll name the attributes A_1, A_2,..., A_n *physical attributes*. Some of them will be used for the logical identification of those entities, some for the representation of the characteristics of the entities, some for referencing other entities, and another attribute might be used for referring the entities themselves.

It is possible to use more attributes, within the user-interface, which are interesting for the user in the context of the system-element. These attributes, together with the physical attributes are referred, in this article, as *presentation attributes*. They can be obtained, using *queries* expressed in SQL, which may imply more relations from the database, connected by join operators.

The tabular image from the user-interface may have as data source, the associated data base relation of that system-element, or a query object. These things, will be also specified, using the system-element description language, SEDL. SEDL enables physical and presentation attributes description for every system-element, the keys, the attributes data type, the attributes properties, etc.

5 The Automatic Generation of the Components

To each system-element, corresponds a component of the final system. The automatic generation process goes from the predefined model of the user-interface of the system-element and the description of the system-element in SEDL. After the generation process there are created one or more classes which implement the user-interface for that system-element.

The author developed a prototype of the generation process, using the VBA language and the Microsoft Access. The generated components are MS Access objects, which the developer may view and even add new functionality or modify the contents of the generated objects. The main gain of the automatic generation is that the developer find a lot of work already done, in general a routine work and theoretical, without errors. This is another gain, that the automatic generated components do not require testing.

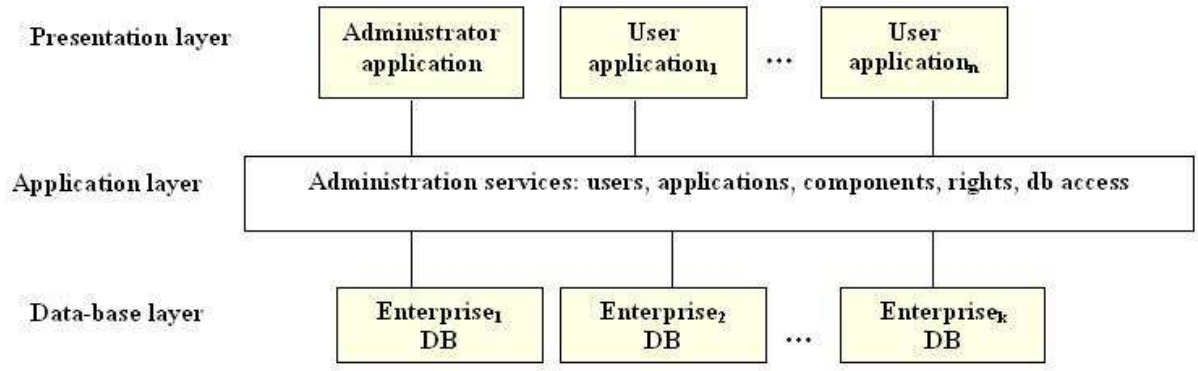


Figure 3: A predefined architecture for an EIS

6 Using a predefined architecture for the EIS

The EIS resulting from the developing process, uses a predefined architecture organized on three layers, as in figure 3. The architecture presented in figure 3, is the architecture of a multi-user, multi-application and multi-enterprise information system. The application layer and the administrator application are predefined. The user

applications, has a predefined kernel, to which we add a number of automatic generated components. The data bases may be also automatic generated or created by the developer. The author also developed a prototype for this architecture.

7 Conclusions

The article presents an EIS development process based on the automatic generation of components. It combine the advantage of using analyze and design patterns for the domain concepts, with the automatic generation of components.

The main advantage of this process consists in minimizing the effort of the components code generation and testing, the developer being more concerned on identifying and specifying the system concepts.

The changes of the system requirements may be done more easily using this approach, because the changes affect only the specification of some concepts, the corresponding components being generated again by executing the automatic code generator.

Another advantage of this approach must be considered when the development tools are changed. We must rewrite only the automatic code generator for each concept type and the predefined components of the system architecture.

This approach is best for developing a customized EIS, rather than a general and very complex EIS.

There are also some inconveniences of this process. The patterns design, the language for the concepts specification, the design and implementation of the automatic code generator are tasks that require high competence and effort for their achievement.

References

[1] M. Fowler, K. Scott, *UML Distilled, Second Edition A brief Guide to the Standard Object Modeling Language*, Addison Wesley, 1999.

[2] D. Oprea, D. Airinei, M. Fotache, *Sisteme informationale pentru afaceri*, Polirom, 2002.

[3] R. Pressman, *Software Engineering, A Practitioner's Approach*, McGrow-Hill, 2000.

[4] D. Zaharie, I. Rosca, *The Objectual Design of Information Systems*, Dual Tech, 2003.

Adrian Alexandrescu
Ovidius University of Constanta
Numerical Methods and Computer Science Department
124 Mamaia Blvd., Constanta, Romania
E-mail: aalexandrescu@univ-ovidius.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 173-178

An Integral Geometry Problem on Three Dimensional Space

A.Kh. Amirov, M. Yildiz

Abstract: In this work, existence, uniqueness and stability of solution of an inverse problem of finding right hand side of the equation $L_1 u \equiv v\nabla_x u = \lambda$ in region Ω_2 is investigated. Here:

$$L_1 u \equiv v\nabla_x u = u_{x_1}\cos\varphi_1 + u_{x_2}\sin\varphi_1\cos\varphi_2 + u_{x_3}\sin\varphi_1\sin\varphi_2 = \lambda,$$

$$\Omega_2 = \left\{(x,v) : x \in D, v \in S^2\right\}$$

D is a bounded region of $\mathbb{R}^3$, S^2 is a unit sphere and

$$v = \left\{(\cos\varphi_1, \sin\varphi_1\cos\varphi_2, \sin\varphi_1\sin\varphi_2)\right\}, \varphi = (\varphi_1,\varphi_2), \varphi_1 \in (0,\pi), \varphi_2 \in (0,2\pi).$$

This problem is investigated by using method of Galerkin. Integral geometry and inverse problems in practice can be applied to the following areas; Medicine, technical tomography, seismology, ecology, etc.
Keywords: Integral geometry, inverse problem, method of Galerkin

1 Introduction

In this paper, three-dimensional inverse problem for the transport equation is studied. Similar problems were investigated in [1-3].

Let S^2 be unit sphere, and let D be a sufficiently smooth domain of the space $\mathbb{R}^3$. Let

$$\Omega_2 = \left\{(x,v) : x \in D,\ v \in S^2\right\},\ v = (\cos\varphi_1, \sin\varphi_1\cos\varphi_2, \sin\varphi_1\sin\varphi_2)$$

$$\frac{\partial}{\partial\theta_1} = \frac{\partial}{\partial\varphi_1},\ \frac{\partial}{\partial\theta_2} = \frac{1}{\sin\varphi_1}\frac{\partial}{\partial\varphi_2},\ \varphi_1 \in (0,\pi),\ \varphi_2 \in (0,2\pi)$$

$$\Gamma_2 = \partial D \times (0,\pi) \times (0,2\pi),\ \varphi = (\varphi_1,\varphi_2).$$

We denote by $\mathcal{M}$, the set of all finite linear combinations of the following functions,

$$w(x)P_{n,m}(\cos\varphi_1)\sin m\varphi_2, w(x)P_{n,m}(\cos\varphi_1)\cos m\varphi_2$$

where $n,m = 0,1,2,\cdots, P_{n,0}(x) \equiv P_n(x)$, $P_n(x)$ are the Legendre polynomials $P_{n,m}(x) = (1-x^2)^{m/2}\frac{d^m}{dx^m}P_n(x)$, and $w(x)$ is an arbitrary real function from $C^2(D)$.

2 Three-Dimensional Inverse Problem

$\overline{H}^1(\Omega_2)$ is the space of real functions obtained through the closure of the set $\mathcal{M}$ in the norm $\|u\|_1^- = \left[(u,u)_1^-\right]^{1/2}$, where

$$(u,z)_1^- = \int\limits_{\Omega_2}\left(uz + \sum_{i=1}^{3} u_{x_i}z_{x_i} + \sum_{i=1}^{2} u_{\theta_i}z_{\theta_i}\right)d\Omega_2,\ d\Omega_2 = dxdS^2$$

dx is the volume element of the domain D and dS^2 is the area element of the sphere S^2. $\overline{H}^{1,2}(\Omega_2)$ is the real Hilbert space obtained by completion of the set $\mathcal{M}$ in the norm

$$(u,z)_{1,2}^{\sim} = \int\limits_{\Omega_2}\left(\sum_{i=1}^{3}\sum_{j=1}^{2} z_{x_i\theta_j}u_{x_i\theta_j}\right)d\Omega_2.$$

In the domain Ω_2 we consider the equation

$$L_1 u \equiv v\nabla_x u = \lambda \tag{1}$$

with right-hand side λ satisfying the following condition: For an arbitrary function $\eta \in \overline{H}^{1,2}(\Omega_2)$, vanishing on Γ_2, the equation

$$\int\limits_{\Omega_2} \lambda \left[\frac{\partial}{\partial \theta_1}(\nabla_x \eta \cdot v') + (\nabla_x \eta \cdot v') \cot \varphi_1 + \frac{\partial}{\partial \theta_2}(\nabla_x \eta \cdot v'') \right] d\Omega_2 = 0 \tag{2}$$

is satisfied. Here $v' = (-\sin\varphi_1, \cos\varphi_1 \cos\varphi_2, \cos\varphi_1 \sin\varphi_2)$, $v'' = (0, -\sin\varphi_2, \cos\varphi_2)$ and $(v \cdot \nabla_x u)$ is the scalar product in $\mathbb{R}^3$.

2.1 Problem

If given $u|_{\Gamma_2} = u_2(x,v)$, it is necessary to determine the pair of functions (u,λ) from equation (1) satisfying condition (2).

The function λ, having the form $\lambda = g(x) + \psi(\varphi)$, where $g, \psi \in C^1$, satisfies relation (2) for an arbitrary function $\eta \in \overline{H}^{1,2}(\Omega_2)$ such that $\eta = 0$ on Γ_2. Now we start to solve the equation (2). Let the function $u(x,v)$ be twice continuously differentiable on Ω_2, and$L_2 = 2(\nabla_x u \cdot v') \frac{\partial}{\partial \theta_1} + 2(\nabla_x u \cdot v'') \frac{\partial}{\partial \theta_2}$. After some calculations we get$L_2 L_1 u = I(u) + D(u)$ where

$$I(u) = |\nabla_x u|^2 + \left(u_{x_1} \cos\varphi_1 + u_{x_2} \sin\varphi_1 \cos\varphi_2 + u_{x_3} \sin\varphi_1 \sin\varphi_2 \right)^2$$

and

$$D(u) = \frac{\partial}{\partial x_1} \left[u_{x_2} u_{\varphi_1} \cos\varphi_2 + u_{x_3} u_{\varphi_1} \sin\varphi_2 - u_{x_2} u_{\varphi_2} \cot\varphi_1 \sin\varphi_2 + \right.$$

$$\left. + u_{x_3} u_{\varphi_2} \cot\varphi_1 \cos\varphi_2 \right] + \frac{\partial}{\partial x_2} \left[-u_{x_1} u_{\varphi_1} \cos\varphi_2 + u_{x_1} u_{\varphi_2} \frac{\cos\varphi_2}{\sin\varphi_1} \sin\varphi_2 + u_{x_3} u_{\varphi_2} \right] +$$

$$+ \frac{\partial}{\partial x_3} \left[-u_{x_1} u_{\varphi_1} \sin\varphi_2 - u_{x_1} u_{\varphi_2} \cot\varphi_1 \cos\varphi_2 - u_{x_2} u_{\varphi_2} \right] +$$

$$+ \frac{1}{\sin\varphi_1} \frac{\partial}{\partial \varphi_1} \left[-u_{x_1}^2 \cos\varphi_1 \sin^2\varphi_1 + u_{x_1} u_{x_2} \sin\varphi_1 \cos\varphi_2 \cos 2\varphi_1 + \right.$$

$$+ u_{x_1} u_{x_3} \sin\varphi_1 \cos 2\varphi_1 \sin\varphi_2 + u_{x_2}^2 \sin^2\varphi_1 \cos\varphi_1 \cos^2\varphi_2 +$$

$$\left. + u_{x_1} u_{x_3} \sin^2\varphi_1 \cos\varphi_1 \sin 2\varphi_2 + u_{x_3}^2 \sin^2\varphi_1 \cos\varphi_1 \sin^2\varphi_2 \right] + \frac{\partial}{\partial \varphi_2} \left[-u_{x_1} u_{x_2} \cot\varphi_1 \sin\varphi_2 + \right.$$

$$\left. + u_{x_3} u_{x_2} \cos 2\varphi_2 - u_{x_2}^2 \sin\varphi_2 \cos\varphi_2 + u_{x_1} u_{x_3} \cot\varphi_1 \cos\varphi_2 + u_{x_3}^2 \sin\varphi_2 \cos\varphi_2 \right].$$

Here we transform each term in the expression $L_2 L_1 u$ in such a way as to obtain an expression containing only $u_{x_i} u_{x_j}$, $u_{x_i} u_{\varphi_k}$ and the divergent terms, where $i, j = 1,2,3$ and $k = 1,2$. Firstly we prove uniqueness of the solution of the problem. Let $u(x,\varphi)$ be a solution of the problem such that $u \in \overline{H}^{1,2}(\Omega)$, $u = 0$ on Γ_2.

We multiply equation (1) scalarly by $\widetilde{L}_2 u$ in the space $L_2(\Omega_2)$, where

$$\widetilde{L}_2 u = \frac{\partial}{\partial \theta_1} (\nabla_x u \cdot v') + (\nabla_x u \cdot v') \cot\varphi_1 + \frac{\partial}{\partial \theta_2} (\nabla_x u \cdot v'').$$

Then, taking equation (2) into account we have, $\left(L_1 u, \widetilde{L}_2 u \right)_{L_2(\Omega_2)} = 0$. Let the sequence $\{u_n\} \subset \mathcal{M}$ tend to $u(x,\varphi)$ in the norm of $\overline{H}^{1,2}(\Omega_2)$, (as $n \to \infty$) and let $u_n = 0$ on Γ_2. From the continuity of the scalar product of the space $L_2(\Omega_2)$ it follows that, as $n \to \infty$,we have expression three

$$\left(L_1 u_n, \widetilde{L}_2 u_n \right)_{L_2(\Omega_2)} \to \left(L_1 u, \widetilde{L}_2 u \right)_{L_2(\Omega_2)}. \tag{3}$$

We now transform the expressions $\left(L_1 u_n, \widetilde{L}_2 u_n \right)_{L_2(\Omega_2)}$ as follows:

$$-2 \left(L_1 u_n, \widetilde{L}_2 u_n \right)_{L_2(\Omega_2)} \equiv -2 \int\limits_{\Omega_2} L_1 u_n \left[\frac{\partial}{\partial \theta_1} (\nabla_x u_n \cdot v') + \right.$$

$$+\left(\nabla_x u_n \cdot v'\right)\cot\varphi_1 + \tfrac{\partial}{\partial\theta_2}\left(\nabla_x u_n \cdot v''\right)\Big] d\Omega_2 = \tag{4}$$

$$= 2\int\limits_{\Omega_2}\left[\tfrac{\partial}{\partial\theta_1}\left(L_1 u_n\right)\left(\nabla_x u_n \cdot v'\right) + \left(\nabla_x u_n \cdot v''\right)\tfrac{\partial}{\partial\theta_2}\left(L_1 u_n\right)\right] d\Omega_2$$

since $u_n \in \mathcal{M}$, $d\Omega_2 = \sin\varphi_1 dx d\varphi_1 d\varphi_2$ and

$$L_1 u_n \frac{\partial}{\partial\theta_1}\left(\nabla_x u_n \cdot v'\right) + L_1 u_n \left(\nabla_x u_n \cdot v'\right)\cot\varphi_1 = \frac{1}{\sin\varphi_1}\left[L_1 u_n \frac{\partial}{\partial\varphi_1}\left(\nabla_x u_n \cdot v'\right)\sin\varphi_1 +\right.$$

$$\left. + L_1 u_n \left(\nabla_x u_n \cdot v'\right)\cos\varphi_1\right] = \frac{1}{\sin\varphi_1}\left\{\frac{\partial}{\partial\varphi_1}\left[L_1 u_n \left(\nabla_x u_n \cdot v'\right)\sin\varphi_1\right]\right\} -$$

$$- \frac{\partial}{\partial\varphi_1}\left(L_1 u_n\right)\left(\nabla_x u_n \cdot v'\right)\sin\varphi_1.$$

Noting that $u_n(x,v)$ is specially dependent on φ_1, φ_2, $u_n(x_1\cos\varphi_1, \sin\varphi_1\cos\varphi_2, \sin\varphi_1\sin\varphi_2)$ it is easy to see in equality (4) that the integral on the right-hand side has meaning and that the transfer of the derivatives with respect to θ_1, θ_2 by $L_1 u_n$ is valid.

We obtain from equality (4)

$$-2\left(L_1 u_n, \widetilde{L}_2 u_n\right)_{L_2(\Omega_2)} = \int\limits_{\Omega_2}\left[I\left(u_n\right) + D\left(u_n\right)\right] d\Omega_2 = \int\limits_{\Omega_2} I\left(u_n\right) d\Omega_2. \tag{5}$$

The divergent portion $D(u_n)$ during the integration over the domain Ω_2, drops out on account of the 2π-periodicity of the function u_n with respect to φ_2 and the fact that $u_n = 0$ on Γ_2 and also on account of the presence in all the terms subjected to the $\partial/\partial\varphi_1$ sign of the factor $\sin\varphi_1$,except for several terms containing derivatives with respect to φ_2. From the special dependence of the function u_n on Γ it follows that these terms also have the factor $\sin\varphi_1$.

From relations (3) and (5), we have equality

$$\int\limits_{\Omega_2} I\left(u\right) d\Omega_2 = 0.$$

The positive-definiteness of the quadratic form $I(u)$, the condition $u = 0$ on Γ_2 and the last equality show that $u = 0$ on Ω_2. Then from equation (1) we find also that $\lambda = 0$ on Ω_2. This completes the proof of the uniqueness of solution to the problem.

Now we prove the existence of the solution of the problem by Galerkin method. Let the domain $D \subset \mathbb{R}^3$ has the form $D = \left\{x \in \mathbb{R}^3 : x_1 \in (a,b),\ x' = (x_2, x_3) \in D'\right\}$, where $D' \subset \mathbb{R}^2$ is a domain with a sufficiently smooth boundary $\partial D'$, and a, b are arbitrary numbers with $a < b$ and $\Gamma_2' = (a,b) \times \partial D' \times (0,\pi) \times (0, 2\pi)$.

We denote by $\overline{H}_1^0(\Omega_2)$, $\overline{H}_{1,2}^0(\Omega_2)$ the spaces of the function obtained through the closure of the set $\overline{\mathcal{M}}$ in the norms of the sets $\overline{H}^1(\Omega_2)$ and $\overline{H}^{1,2}(\Omega_2)$ respectively. The set $\overline{\mathcal{M}}$ is the linear hull of function of the form $w_i(x')P_{n,m}(\cos\varphi_1)\sin m\varphi_2$; $w_i P_{n,m}(\cos\varphi_1)\cos m\varphi_2$, $P_{n,0}(x) \equiv P_n(x)$, where $n, m, i = 0, 1, 2, \cdots$; the functions $w_i(x')$, twice continuously differentiable, are finite linear combinations everywhere dense on $\overset{0}{H}{}^1(D')$, linearly independent, $w_i = 0$ on $\partial D'$.

To prove the existence of the solution of the problem, in addition to the conditions above, we assume that $u_2 \in C(\Gamma_2)$ and $u_2(x,v)$ are three times continuously differentiable functions of their arguments. Then we find a pair of functions (u, λ) such that $u \in \overline{H}^1(\Omega_2)$ and $\lambda \in L_2(\Omega_2)$.

Since $\partial D' \in C^3$ and $u_2(x,v) \in C^3(\Gamma_2')$, there exists a three times differentiable functions $w(x,v)$ on Ω_2 such that $w = u_2$ on Γ_2'.

We introduce the function $u_1 = u - w$. If we redesignate u_1 by u, the function u will then satisfy the equation

$$L_1 u = \lambda + F \tag{6}$$

and the conditions

$$u|_{x_1=a} = \overline{\varphi}_1(x', v),\ \ u|_{x_1=b} = \overline{\varphi}_2(x', v),\ \ u|_{\Gamma_2'} = 0 \tag{7}$$

where $\overline{F} = L_1 w$, $\overline{\varphi}_1 = (u_2 - w)|_{x_1=a}$, $\overline{\varphi}_2 = (u_2 - w)|_{x_1=b}$.

For simplicity in what follows, we assume that the functions $\overline{\varphi}_1,\ \overline{\varphi}_2$ are identically equal to zero. The case in which these functions are nonzero does not entail additional complications. The conditions (6) can then be written in the form $u|_{\Gamma_2}=0$.

We now consider the following auxiliary problem. Find the solution of the equation

$$L_3u\equiv\varepsilon\frac{\partial^3u}{\partial x_1^3}+\left(-\sin\varphi_2\frac{\partial}{\partial x_2}+\cos\varphi_2\frac{\partial}{\partial x_3}\right)\frac{\partial}{\partial\theta_2}(L_1u)+$$

$$+\left(-\sin\varphi_1\frac{\partial}{\partial x_1}+\cos\varphi_1\cos\varphi_2\frac{\partial}{\partial x_2}+\cos\varphi_1\sin\varphi_2\frac{\partial}{\partial x_3}\right)\frac{\partial}{\partial\theta_1}(L_1u)=\overline{F} \tag{8}$$

satisfying the conditions

$$u|_{\Gamma_2}=0\ ,\ u_{x_1}|_{x_1=a}=0 \tag{9}$$

where $\overline{F}=v'\cdot\nabla_x\overline{F}_{\theta_1}+v''\cdot\nabla_x\overline{F}_{\theta_2}$ and $\varepsilon>0$.

We seek an approximate solution of problem (8)-(9) in the form

$$u_{\varepsilon N}=\sum_{i,n=1}^{N}\left\{a_{0,i}^{(n)}(x_1,\varepsilon)P_n(\cos\varphi_1)+\sum_{m=1}^{n}\left[a_{m,i}^{(n)}(x_1,\varepsilon)\cos m\varphi_2+\right.\right.$$
$$\left.\left.+\,b_{m,i}^{(n)}(x_1,\varepsilon)\sin m\varphi_2\right]P_{n,m}(\cos\varphi_1)\right\}w_i(x').$$

We seek the functions $a_{0,i^{(n)}},a_{m,i^{(n)}},b_{m,i^{(n)}}$ where $i,n=1,2,3,\cdots,N\ ;\ m=1,2,3,\cdots,n$, satisfying the conditions

$$a_{0,i}^{(n)}(a,\varepsilon)=a_{0,i}^{(n)}(b,\varepsilon)=\frac{d}{dx_1}a_{0,i}^{(n)}(a,\varepsilon)=0,$$

$$a_{m,i}^{(n)}(a,\varepsilon)=a_{m,i}^{(n)}(b,\varepsilon)=\frac{d}{dx_1}a_{m,i}^{(n)}(a,\varepsilon)=0,\ b_{m,i}^{(n)}(a,\varepsilon)=b_{m,i}^{(n)}(b,\varepsilon)=\frac{d}{dx_1}b_{m,i}^{(n)}(a,\varepsilon)=0, \tag{8′}$$

from the following relations

$$\left(L_3u_{N\varepsilon}-\overline{F},w_iP_n(\cos\varphi_1)\right)_{L_2(D'\times S^2)}=0, \tag{10}$$

$$\left(L_3u_{N\varepsilon}-\overline{F},w_iP_{n,m}(\cos\varphi_1)\cos m\varphi_2\right)_{L_2(D'\times S^2)}=0, \tag{11}$$

$$\left(L_3u_{N\varepsilon}-\overline{F},w_iP_{n,m}(\cos\varphi_1)\sin m\varphi_2\right)_{L_2(D'\times S^2)}=0, \tag{12}$$

where $i,n=1,2,3,\cdots,N$; $m=1,2,3,\cdots,n$.

To simplify the writing we denote the vector of unknown functions $\left(a_{m,i^{(n)}},b_{m,i^{(n)}}\right)$ where $i,n=0,1,2,\cdots,N$, $m=0,1,2,\cdots,n$, by the vector $(u_i(x_1,\varepsilon))$, $i=1,2,\cdots,l(N)$, $l(N)=N^2+\sum_{n=1}^{N}n(n+1)$.

Since for an arbitrary positive integer N the system of functions

$$\mathcal{M}_N=\{w_iP_{n,m}(\cos\varphi_1)\sin m\varphi_2,w_iP_{n,m}(\cos\varphi_1)\cos m\varphi_2,$$
$$i,n=0,1,\cdots,N,\ m=0,1,\cdots,n\}$$

is linearly independent, the determinant of the matrix $\left((r_i,r_j)_{L_2(D'\times S^2)}\right)$ is different from zero, where r_i and r_j are arbitrary elements of the set $\mathcal{M}_N$. We can therefore solve the system of ordinary differential equations (10)-(12) for the higher derivatives

$$\varepsilon\frac{d^3u_i}{dx_1^3}+\sum_{j=1}^{l(N)}\left[a_{ij}(x)\frac{d^2}{dx_1^2}u_j+b_{ij}(x_1)\frac{d}{dx_1}u_j+c_{ij}(x_1)u_j\right]=F_i(x_1),\ i=1,2,\cdots,l(N) \tag{13}$$

Here a_{ij},b_{ij},c_{ij},F_i are known functions, continuous on $[a,b]$. From equations (8') we have

$$u_i(a,\varepsilon)=u_i(b,\varepsilon)=\tfrac{d}{dx_1}u_i(a,\varepsilon)=0,\ i=1,2,\cdots,l(N) \tag{14}$$

Consequently, we arrive at problem (13)-(14): Determine the functions $\{u_i(x,\varepsilon),\ i=1,2,\cdots,l(N)\}$, from the system (13) which satisfy the conditions (14).

We show now that, considering above conditions, problem (13)-(14) possesses a unique solution $\{u_i(x,\varepsilon)\}$, $i=1,2,\cdots,l(N)$. For this purpose it is sufficient to prove that homogeneous system (13)-(14) has only null solution. Suppose, this homogeneous problem (13)-(14) has only non null solution $\overline{u}_N=\{\overline{u}_i(x,\varepsilon)\}$, $i=1,2,\cdots,l(N)$, $\left(\overline{a}_{0,i^{(n)}},\overline{a}_{m,i^{(n)}},\overline{b}_{m,i^{(n)}}\right)$. Since $\overline{u}_N$ is a solution of homogeneous problem (13)-(14), the vector $\left(\overline{a}_{0,i^{(n)}},\overline{a}_{m,i^{(n)}},\overline{b}_{m,i^{(n)}}\right)$ satisfies system (10)-(12) for $\overline{F}=0$ and for conditions (8′). This can be seen by multiplying the homogeneous system (13) by the matrix $\left((r_i,r_j)_{L_2(D'\times S^2)}\right)$, defined above. In the system of equations (10)-(12) with $\overline{F}=0$, we replace $\left(a_{0,i^{(n)}},a_{m,i^{(n)}},b_{m,i^{(n)}}\right)$ by $\overline{a}_{0,i^{(n)}},\overline{a}_{m,i^{(n)}},\overline{b}_{m,i^{(n)}}$ respectively and we multiply the $(i,n)-$th equation of the system (10) by $-2\overline{a}_{0,i}^{(n)}$, the $(t,n,m)-$th equation of system (11) by $-2\overline{a}_{m,i}^{(n)}$, the $(i,n,m)-$th equation of system (12) by $-2\overline{b}_{m,i}^{(n)}$ and then sum over (i,n) from 1 to N, and over m from 1 to n. As a result,

$$-2\left(L_3\overline{u}_{N\varepsilon},\overline{u}_{N\varepsilon}\right)_{L_2(D'\times S^2)}=0 \tag{15}$$

where $\overline{u}_{N\varepsilon}$ denotes the function obtained by replacing, in the expression for $u_{N\varepsilon}$, the quantities $a_{0,i^{(n)}}$, $a_{m,i^{(n)}}$, $b_{m,i^{(n)}}$ by $\overline{a}_{0,i^{(n)}},\overline{a}_{m,i^{(n)}},\overline{b}_{m,i^{(n)}}$ respectively. We can verify directly that

$$-2\varepsilon\frac{\partial^3}{\partial x_1^3}\overline{u}_{\varepsilon N}\overline{u}_{\varepsilon N}=-2\varepsilon\left(\frac{\partial^2}{\partial x_1^2}\overline{u}_{\varepsilon N}\overline{u}_{\varepsilon N}\right)_{x_1}+\varepsilon\frac{\partial}{\partial x_1}\left(\overline{u}_{\varepsilon N x_1}^2\right);$$

$$\begin{aligned}
&-2\left[\left(-\sin\varphi_2\frac{\partial}{\partial x_2}+\cos\varphi_2\frac{\partial}{\partial x_3}\right)\frac{\partial}{\partial\theta_2}(L_1\overline{u}_{\varepsilon N})+\right.\\
&\left.+\left(-\sin\varphi_1\frac{\partial}{\partial x_1}+\cos\varphi_1\cos\varphi_2\frac{\partial}{\partial x_2}+\cos\varphi_1\sin\varphi_2\frac{\partial}{\partial x_3}\right)\frac{\partial}{\partial\theta_1}\right]\times(L_1\overline{u}_{\varepsilon N})\,\overline{u}_{\varepsilon N}\\
&=2\left[\overline{u}_{\varepsilon N}\frac{\partial}{\partial\theta_2}(L_1\overline{u}_{\varepsilon N})\sin\varphi_2\right]_{x_2}-2\left[\overline{u}_{\varepsilon N}\frac{\partial}{\partial\theta_2}(L_1\overline{u}_{\varepsilon N})\cos\varphi_2\right]_{x_3}\\
&+2\left[\overline{u}_{\varepsilon N}\frac{\partial}{\partial\theta_1}(L_1\overline{u}_{\varepsilon N})\sin\varphi_1\right]_{x_1}-2\left[\overline{u}_{\varepsilon N}\frac{\partial}{\partial\theta_1}(L_1\overline{u}_{\varepsilon N})\right]_{x_2}\cos\varphi_1\cos\varphi_2-\\
&-2\left[\overline{u}_{\varepsilon N}\frac{\partial}{\partial\theta_1}(L_1\overline{u}_{\varepsilon N})\cos\varphi_1\sin\varphi_2\right]_{x_3}+2\tfrac{\partial}{\partial\theta_2}(L_1\overline{u}_{\varepsilon N})\left(\nabla_x\overline{u}_{\varepsilon N}\cdot\nu''\right)+\\
&+2\tfrac{\partial}{\partial\theta_1}(L_1\overline{u}_{\varepsilon N})\left(\nabla_x\overline{u}_{\varepsilon N}\cdot\nu'\right).
\end{aligned}\tag{16}$$

Then $\overline{u}_{\varepsilon N}=0$ on Γ_2, and then integrating the equation (15) over $[a,b]$ with equation (16) taken into account, we obtain

$$\int_{\Omega_2}I(\overline{u}_{\varepsilon N})d\Omega_2=0. \tag{17}$$

However, we have $I(\overline{u}_{\varepsilon N})>\alpha_1\left(|\nabla_x\overline{u}_{\varepsilon N}|^2+|\nabla_\theta\overline{u}_{\varepsilon N}|^2\right)$. Then from equation (17), $\overline{u}_{\varepsilon N}=0$ on Γ_2.

Since the system $\mathcal{M}_N$ is linearly independent, then $\overline{a}_{0,i^{(n)}}=0$, $\overline{a}_{m,i^{(n)}}=0$, $\overline{b}_{m,i^{(n)}}=0$, $i,n=1,2,\cdots,N$, $m=1,2,\cdots,m$. But this contradicts our assumption. This contradiction shows that the homegeneous problem (13)-(14) has only null solution. Consequently, problem (8′)-(12) has for an arbitrary continuous right member $\overline{F}$ a unique three-times continuously differentiable solution.

We estimate, in terms of $\overline{F}$, the right-hand side of equation (10), where $u_{\varepsilon N}$ is determined with the aid of the coefficients $a_{0,i^{(n)}},a_{m,i^{(n)}},b_{m,i^{(n)}}$, $i,n=1,2,\cdots,N$ by the solution of problem (8′)-(12). To do this, we multiply the (i,n)-th equation of system (10) by $-2a_{0,i^{(n)}}$, (i,n,m)-th equation of system (11) by $-2a_{m,i^{(n)}}$, $(i,n)-$th equation of the system (12) by $-2b_{m,i^{(n)}}$, and then sum over (i,n) from 1 to N and over m from 1 to n. Then we get

$$-2\left(L_3u_{\varepsilon N},u_{\varepsilon N}\right)_{L_2(D'\times S^2)}=-\left(\overline{F},u_{\varepsilon N}\right)_{L_2(D'\times S^2)} \tag{18}$$

If we take into account the fact that $\overline{F}=\nu'\cdot\nabla_x\,\overline{F}_{\theta_1}+\nu''\cdot\nabla_x\,\overline{F}_{\theta_2}$, $u_{\varepsilon N}=0$ on Γ_2, then integrating equation (18)

over $[a,b]$, we can estimate the right side of the resulting equation as follows:

$$-2\int_{\Omega_2}\overline{F}u_{\varepsilon N}d\Omega_2=\int_{\Omega_2}\left[\left(v'\cdot\nabla_x u_{\varepsilon N}\right)\overline{F}_{\theta_1}+\left(v''\cdot\nabla_x u_{\varepsilon N}\right)\overline{F}_{\theta_2}\right]d\Omega_2\leq$$

$$\leq\beta\int_{\Omega_2}\left|\nabla_\theta\overline{F}\right|^2d\Omega_2+\frac{1}{\beta}\int_{\Omega_2}\left|\nabla_x u_{\varepsilon N}\right|^2d\Omega_2$$

where $\beta^{-1}<\alpha_1$. In obtaining the last inequality we have used the Cauchy-Bunyakovskii inequality.

As was shown above, the left side of equation (18), integrated over $[a,b]$, is equal to $\int_{\Omega_2}I(u_{\varepsilon N})d\Omega_2$. Consequently from equation (18) we have

$$\int_{\Omega_2}\alpha_1\left(\left|\nabla_x u_{\varepsilon N}\right|^2+\left|\nabla_\theta u_{\varepsilon N}\right|^2\right)d\Omega_2\leq\int_{\Omega_2}I(u_{\varepsilon N})d\Omega_2\leq\beta\int_{\Omega_2}\left|\nabla_\theta\overline{F}\right|^2d\Omega_2+\beta^{-1}\int_{\Omega_2}\left|\nabla_x u_{\varepsilon N}\right|^2d\Omega_2.$$

Since the domain Ω_2 is bounded and $u_{\varepsilon N}=0$ on Γ_2, it folows from the last inequality

$$\|u_{\varepsilon N}\|_{\overline{H}_1^0(\Omega_2)}\leq c\left\|\left|\nabla_\theta\overline{F}\right|\right\|_{L_2(\Omega_2)}\tag{19}$$

where $c>0$ independent of N and ε.

We can show using the last inequality that there exists a sequence $u_{\varepsilon_k N}$ such that $\frac{\partial}{\partial x_i}u_{\varepsilon_k N}$, $\frac{\partial}{\partial\theta_j}u_{\varepsilon_k N}$, $i=1,2,3$, $j=1,2$ converge weakly in $L_2(\Omega_2)$ as $N\to\infty$, $k\to\infty$ $(\varepsilon_k\to 0)$ to u_{x_i}, u_{θ_j} respectively, where $u\in\overline{H}_1^0(\Omega_2)$ is the weak limit in $\overline{H}_1^0(\Omega_2)$ of the sequence $u_{\varepsilon_k N}$.

Multiplying equations (10)-(12) by an arbitrary function $q(x_1)\in C^2(a,b)$ we transfer the required derivatives onto the function

$$q(x_1)P_{n,m}(\cos\varphi_1)\cos m\varphi_2,\ qP_{n,m}(\cos\varphi_1)\sin m\varphi_2,\ qP_n(\cos\varphi_1)$$

where $q(a)=q(b)=q_{x_1}(b)=0$.

Then, noting that $u_{\varepsilon N}=0$ on Γ_2, $d\Omega_2=\sin\varphi_1 dDd\varphi_1 d\varphi_2$, and integrating the resulting equations over $[a,b]$ for $N\geq\overline{n}\geq\overline{m}$ and ε_k, we have

$$\varepsilon_k\left(\tfrac{\partial}{\partial x_1}u_{\varepsilon_k}N,\tfrac{\partial^2}{\partial x_1^2}r^{(1)}_{\overline{n},\overline{m}}\right)_{L_2(\Omega_2)}+$$

$$\left(L_1u_{\varepsilon_k}N-\overline{F},\ \left[\tfrac{\partial}{\partial\theta_1}\left(\nabla_x r^{(1)}_{\overline{n},\overline{m}}\cdot v'\right)+\tfrac{\partial}{\partial\theta_2}\left(\nabla_x r^{(1)}_{\overline{n},\overline{m}}\cdot v''\right)+\left(\nabla_x r^{(1)}_{\overline{n},\overline{m}}\cdot v'\right)\cot\varphi_1\right]\right)_{L_2(\Omega_2)}=0$$

where $l=1,2$, $r^{(1)}_{\overline{n},\overline{m}}=qP_{\overline{n},\overline{m}}(\cos\varphi_1)\sin\overline{m}\varphi_2$, $r^{(2)}_{\overline{n},\overline{m}}=qP_{\overline{n},\overline{m}}(\cos\varphi_1)\cos\overline{m}\varphi_2$, $\overline{m}=0,1,2,\cdots,\overline{n}$, $P_{\overline{n},0}=P_{\overline{n}}$.

Since linear combinations of the functions $r^{(l)}_{\overline{n},\overline{m}}$ are everywhere dense on $\overline{H}^0_{1,2}(\Omega_2)$, we have upon passing to the limit in the inequality (19) as $N\to\infty$, $k\to\infty$ $(\varepsilon_k\to 0)$,

$$\left(L_1u-\overline{F},\left[\tfrac{\partial}{\partial\theta_1}\left(\nabla_x\eta\cdot v'\right)+\tfrac{\partial}{\partial\theta_2}\left(\nabla_x\eta\cdot v''\right)+\left(\nabla_x\eta\cdot v'\right)\cot\varphi_1\right]\right)_{L_2(\Omega_2)}=0$$

for arbitrary $\eta\in\overline{H}^0_{1,2}(\Omega_2)$. This completes the proof of the exist problem.

References

[1] M. M. Lavrent'ev, V. G. Romanov, and S. P. Shishatskii, Ill-Posed Problems of Mathematical Physics and Analysis. American Mathematical Society (1986).

[2] M. M. Lavrent'ev and Yu. E. Anikonov, "On a class of problems of integral geometry," Dokl. Akad. Nauk SSSR, 176, No. 5, 1240-1241 (1967).

[3] A. Kh Amirov, Existence And Uniqueness Theorems For The Solution Of An Inverse Problem for The Transport Equation. Siberian Mathematical Society (1986).

A. Kh. Amirov, M. Yildiz
Zonguldak Karaelmas University
Department of Mathematics
E-mail: amirov@karaelmas.edu.tr, mustafayildiz2002@hotmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 179-184

Formation Control of Multi-Robots via Fuzzy Logic Technique

A. Bazoula, M.S. Djouadi, H. Maaref

Abstract: In this paper we address the problem of mobile robot formation control. Indeed, the most work, in this domain, have studied extensively classical control for keeping a formation of mobile robots. In this work, we design an FLC (Fuzzy logic Controller) controller for separation and bearing control (SBC). Indeed, the leader mobile robot is controlled to follow an arbitrary reference path, and the follower mobile robot use the FSBC (Fuzzy Separation and Bearing Control) to keep constant relative distance and constant angle to the leader robot. The efficiency and simplicity of this control law has been proved by simulation on different situations.
Keywords: Autonomous mobile robot, Formation control, Fuzzy logic control, Multiple robots.

1 Introduction

Formation control of multiple autonomous mobile robots and vehicles has been studied extensively over the last decade for both theoretic research and practical applications. Various approaches and strategies have been proposed for the formation control of multiple robots.
However, multi-robot coordination methods can be partitioned into three class approaches: virtual structure approach, behavioral approach and leader follower approach. Each of them has several advantages and weaknesses.
The virtual structure approach treats the entire formation as a single virtual rigid structure [1]-[3]. Desired motion is assigned to the virtual structure as a whole, as a result which will trace out trajectories for each robot in the formation to follow. It is easy to prescribe the behavior of the whole group, and maintain the formation very well during the maneuvers. The main disadvantage of the current virtual structure implementation is the centralization, which leads a single point of failure for the whole system.
By behavior based approach, several desired behaviors are prescribed for each robot, and the final action of each robot is derived by weighting the relative importance of each behavior. Possible behaviors include obstacle avoidance, collision avoidance, goal seeking and formation keeping [3]-[5]. The limitation of behavior-based approach is that it is difficult to analyze mathematically, therefore it is hard to guarantee a precise formation control.
In the leader following approach [7], [8], one of the robots is designated as the leader, with the rest being followers. The follower robots need to position themselves relative to the leader and to maintain a desired relative position with respect to the leader. In order, to prescribe a formation maneuver, we need only to specify the leaders motion and the desired relative positions between the leader and the followers.
When the motion of the leader is known, the desired positions (desired distance and orientation) of the followers relative to the leader can be achieved by local control law on each follower. Therefore, in a certain sense, the formation control problem can be seen as a natural extension of the traditional trajectory-tracking problem.
Among all the approaches to formation control reported in the literature, the leader-following method has been adopted by many researchers [2], [3], [6], [8], [9]. In this method, each robot takes another neighboring robot as a reference point to determine its motion. The referenced robot is called a leader, and the robot following it called a follower. Thus, there are many pairs of leaders and followers and complex formations can be achieved by controlling relative positions of these pairs of robots respectively. This approach is characterized by simplicity, reliability and no need for global knowledge and computation.
In this paper, we will develop a method based on the leader-following approach to investigate formation control problem in a group of nonholonomic mobile robots. For this purpose, we design a new controller based on fuzzy logic to drive a fleet of mobile robots in a leader-follower configuration.
The rest of this paper is organized as follows. First motion modeling is revealed. Then, the method of path following used by the leader robot is exposed. After that, the architecture of the fuzzy controller is described. We conclude the paper by some simulation and results.

2 Mobile Robot Modeling

The mobile robots considered in this study are the popular wheeled mobile robots of unicycle type, shown in Fig. 1. The configuration of the robot denoted by $q = (x, y, \theta)^T \in \Re^3$ in the Earth fixed inertial coordinate system

$X-Y$, where $(x(t),y(t))$ represents the position of the mobile robot by the fixed Cartesian coordinates where t is time, and the angle $\theta(t)$, $(-\frac{\pi}{2} \leq \theta(t) \leq \frac{\pi}{2})$ its orientation relatively to the x-axis. The linear and angular velocities are respectively $v(t)$ and $\omega(t)$. The equation of motion of the mobile robot is given by:

$$\begin{aligned} \dot{x}(t) &= v(t)cos\theta(t) \\ \dot{y}(t) &= v(t)sin\theta(t) \\ \dot{\theta}(t) &= \omega(t) \end{aligned} \tag{1}$$

where $v(t)$ and $\omega(t)$ are considered as the inputs to the mobile robot. Their magnitudes are constrained as follows: $\|v(t)\| < a_{max}$, $\|\dot{v}(t)\| < v_{max}$, $\|\omega(t)\| < \omega_{max}$
where v_{max}, a_{max} and ω_{max} are the maximum admissible values for $v(t)$, $a(t)$ and $\omega(t)$, respectively. Furthermore, from equation (1), the mobile robot behavior is subject to an additional nonholonomic constraint:

$$\dot{x}(t)sin\theta(t) - \dot{y}(t)cos\theta(t) = 0 \tag{2}$$

This constraint means that the robot can not move in the direction of the wheel axis (i.e. y).

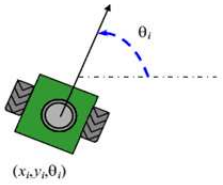


Figure 1: Unicycle Robot

3 Modeling of Leader-Follower Formation

Let us consider the situation shown in Fig.2. A leader and a follower robot are denoted as R_l and R_f, respectively. According to the notation defined in Section II, the states and the inputs of R_l and R_f are newly denoted as (x_l,y_l,θ_l), (x_f,y_f,θ_f), (v_l,ω_l) and (v_f,ω_f) , where the subscript "l" and "f" mean leader and follower, respectively. The equations of motion of both robots are given by (1). The relative distance between the leader and the follower robot is denoted as $d_{lf} = d_{fl}$. The separation bearing angle is ψ_{lf} , they are given by:

$$\begin{aligned} d_{lf} &= \sqrt{(x_l - x_f)^2 + (y_l - y_f)^2} \\ \psi_{lf} &= \pi - arctan2(y_f - y_l, x_f - x_l) - \theta_l \end{aligned} \tag{3}$$

where θ_l is the orientation of the leader and ψ_{lf} is the separation bearing between the leader and the follower.

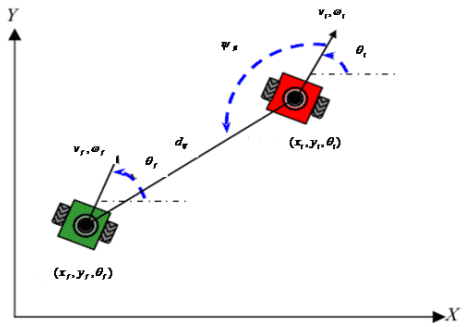


Figure 2: Leader-follower approach for unicycle robot

4 Fuzzy SBC Control

The main goal of the FSBC is to make the follower mobile robot pursue the leader mobile robot. The parameters used to construct the FSBC are shown in Fig.2, where (x_l, y_l) coordinates of center of the leader robot, θ_l represents the orientation angle corresponding to the x-axis. (x_f, y_f) coordinates center of the follower robot, ψ_{lf} is the separation bearing between the the leader and the follower.

$$e(k) = d_{des}(k) - d_{act}(k) \tag{4}$$

$$\Delta e(k) = e(k) - e(k-1) \tag{5}$$

$$\theta(k) = \psi_{lf}(k) - \psi_d(k) \tag{6}$$

$$\Delta\theta(k) = \theta(k) - \theta(k-1) \tag{7}$$

where d_{des} and is the desired distance between leader and the follower robot and ψ_d the desired bearing . We want to design the FSBC for the follower mobile robot such that it can pursue the target mobile robot within a constants distance and bearing. The fuzzy control rules can be represented as a mapping from input linguistic variables e , Δe , θ and $\Delta\theta$ to output linguistic variables v and ω as follows:

$$u(k+1) = FSBC(e(k), \Delta e(k), \theta(k), \Delta\theta(k)) \tag{8}$$

Where u includes the translation and angular velocity. Because translation and angular velocities of the mobile robot can be driven individually, thus we can decompose (8) as follows:

$$v(k+1) = FSBC(e(k), \Delta e(k)) \tag{9}$$

$$\omega(k+1) = FSBC(\theta(k), \Delta\theta(k)) \tag{10}$$

The membership functions of input linguistic variables e, Δe, θ, $\Delta\theta$ and the membership functions of output linguistic variables v and ω are shown in Fig.3 respectively. They are all decomposed into seven fuzzy partitions,

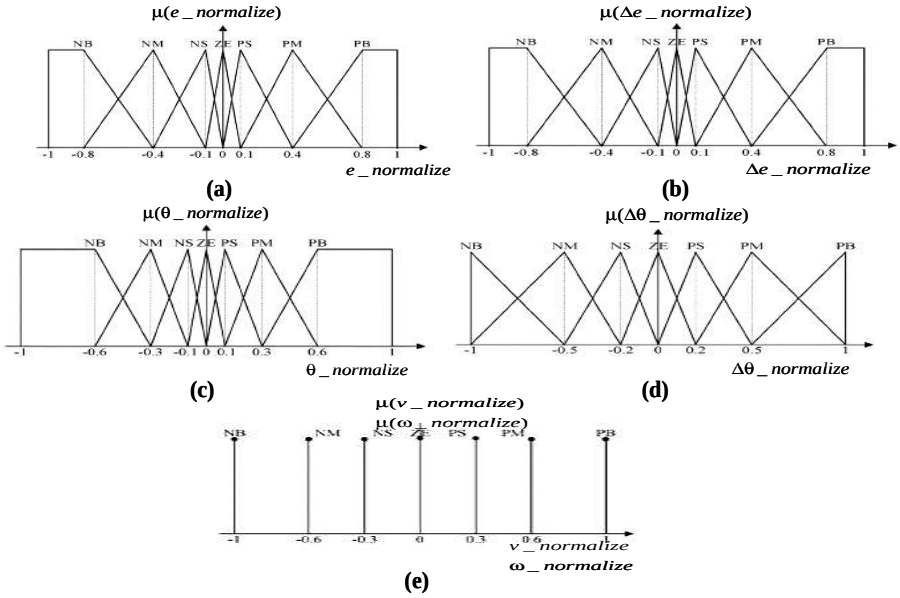


Figure 3: (a) Membership function of e. (b)Membership function of Δe. (c) Membership function of θ. (d) Membership function of $\Delta\theta$. (e) Membership function of v and ω

such as : negative big (**NB**), negative medium (**NM**), negative small (**NS**), zero(**ZE**), positive small **PS**, positive medium (**PM**) and positive big (**PB**).
Since each input is divided into seven fuzzy sets, 49 fuzzy rules for speed control and steering angle control must

E \ dE	NB	NM	NS	ZE	PS	PM	PB
PB	ZE	NS	NM	NB	NB	NB	NB
PM	PS	ZE	NS	NM	NB	NB	NB
PS	PM	PS	ZE	NS	NM	NB	NB
ZE	PB	PM	PS	ZE	NS	NM	NB
NS	PB	PB	PM	PS	ZE	NS	NM
NM	PB	PB	PB	PM	PS	ZE	NS
NB	PB	PB	PB	PB	PM	PS	ZE

Table 1: Rules table of FSBC, (case $u = v$, $E = e$, $dE = \Delta e$, case $u = \omega$, $E = \theta$, $dE = \Delta\theta$)

be determined, respectively. Following the FSMC concept [9]-[11], a diagonal type rule table is adopted (see Table 1). The defuzzification strategy is implemented by the weighted average method.

$$u = \frac{\sum_{j=1}^{49} \mu_j(u_j).u_j}{\sum_{j=1}^{49} \mu_j(u_j)} \tag{11}$$

where u may be the linear velocity or angular velocity command of the follower mobile robot, μ_j is the support of each fuzzy set j, and $\mu_j(u_j)$ is the is the membership function value of each rule.

5 Simulation Results

To illustrate the efficiency of the proposed controller, we simulate a team of 3 nonholonomic mobile robots as shown in figure 4. Initially, the position of the leader (red color) is Rob1(100, 0, 0). The position of the two followers are Rob2(90, 0,0) and Rob3(70, 0,0) (green and blue respectively) the distance inter-robots are initially at $d_{12} = 20$, $d_{23} = 30$, and we want to keep the relative distance and angle between two consecutive robots as a constant value ($d = 20$ and $\theta = 0$).
The simulation results for several different cases are listed as below in Fig.4 to Fig.5. In each figure, the three plots are the simulation results of the robot team evolution, translation velocity of the leader, the angular velocity of the leader, inter-distance error, and angular error form the top to bottom respectively.
Fig.4 shows the case when the reference path is a straight line with a slope of 0.2. The leader goes along a straight line with a constant linear speed (0.2m/s) and constant angular speed (0rad/s). Fig. 4 shows the result when the followers robots keep a constant relative angle ($\theta_d = 0$) with respective to the leader.
Fig.5 shows the case of a reference path, with an initial curvature equal to -0.05 ($\omega = 0.01 rad/s$) and changes later to 0.05 ($\omega = 0.01 rad/s$) at t = 50s.

6 Conclusion

In this paper, we present a controller based on fuzzy logic for the leader following formation of multiple nonholonomic mobile robots. For keeping the distance and bearing between two consecutive mobile robots in desired values, we design two controllers the first for translation velocity its inputs are distance and its derivative, the second one is for angular velocity its inputs are the bearing between the mobile robots orientations and its derivative. Successful simulation have been conducted for various situations and showed the efficiency of the proposed approach.
Authors future work will be focused on the use of this controller to any other formations shape such as "V" shape, Wedge, and experimental verification of presented approach.

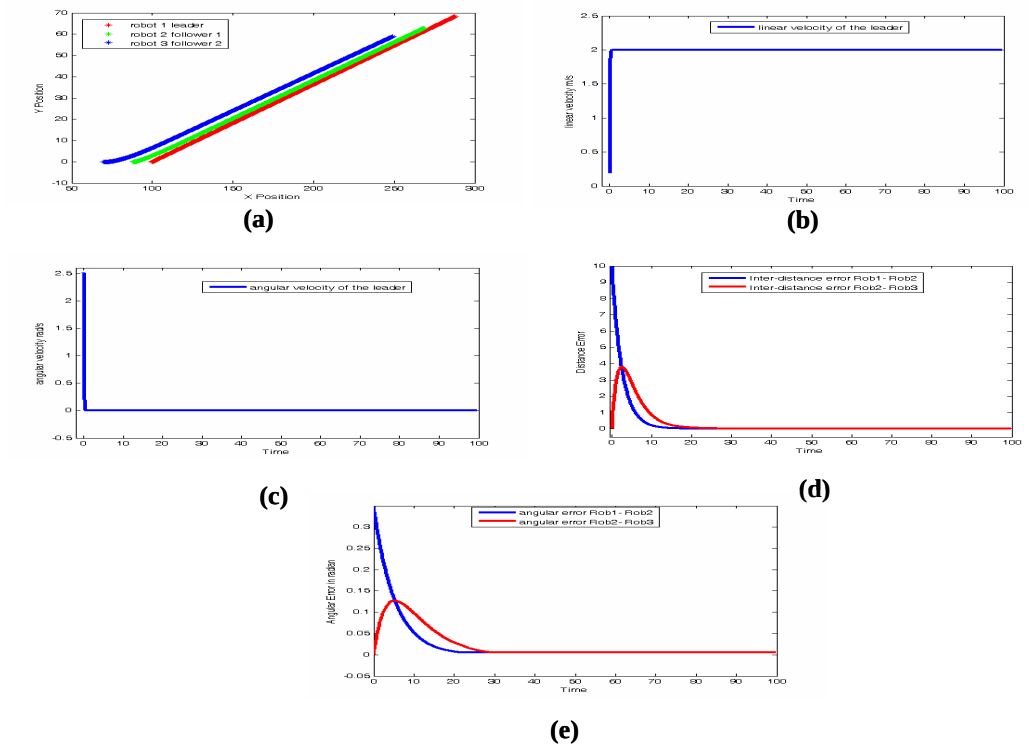


Figure 4: Leader robot goes on straight line and the follower keeps a constant relative distance and angle

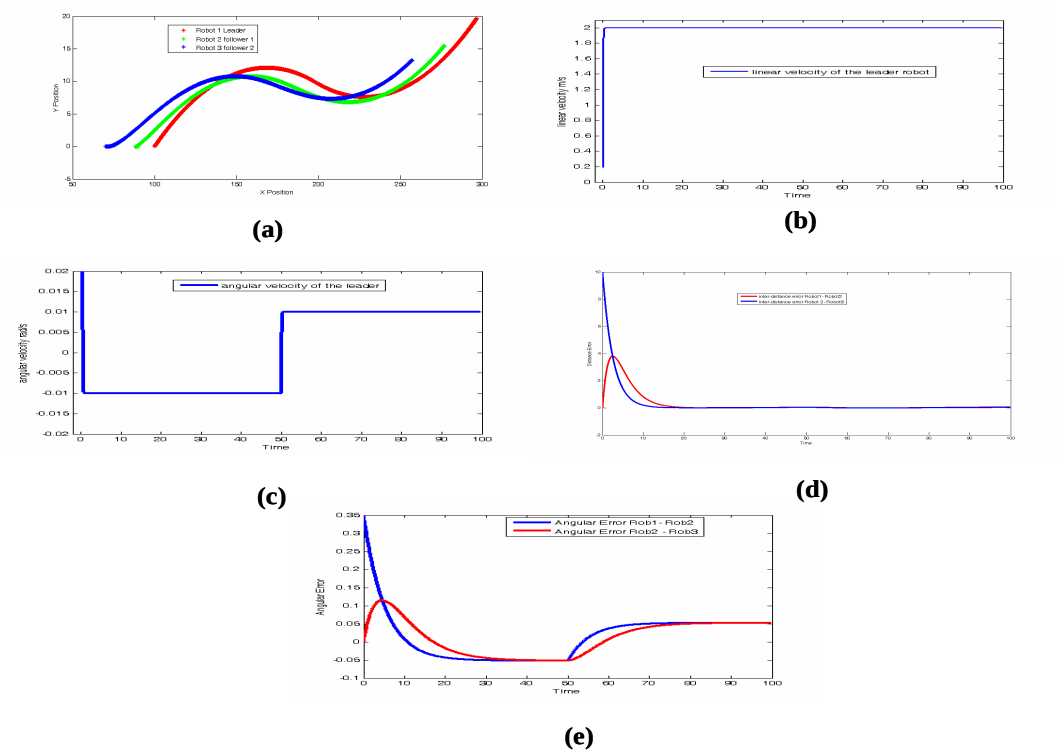


Figure 5: Leader robot goes on reference path that has initially its curvature equal to -0.05 and changes the curvature to 0.05 at time $t = 50s$

References

[1] Kar-Han Tan , M.A.Lewis, "Virtual structures for high precision cooperative control," *Technical Report,* Computer Science Department, University of California, Los Angles, 1997

[2] W. Ren, R. W. Beard, "A Decentralized scheme for spacecraft formation flying via the virtual structure approach," *AIAA Journal of Guidance, Control, and Dynamics,* Revised Submission, June 2003.

[3] M. Egerstedt, K. Hu, "Formation constrained multi-agent control," *proceedings of 2001 IEEE International Conference on Robotics and Automation,* pp .3961-3966, Seoul, Korea, May 21-26, 2001

[4] J.M. Esposito, V. Kumar, "A formalism for parallel composition of reactive and deliberative control objectives for mobile robots," *Technical Report*, VMechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, 2000

[5] H.Yamgachi, "A Cooperative Hunting Behavior by Mobile Robot Troops," *ICRA'98*, pp.3204-3209, Leuven Belgium, May 1998

[6] K. Sugihara and I. Suzuki, "Distributed algorithms for formation of geometric patterns with many mobile robots," *Journal of Robot Systems*, Vol. 9, pp.777-790, 2001.

[7] Jose Sanchez and Rafael Fierro, "Sliding mode control for robot formations," *Proc. IEEE Int. Symposium on Intelligent Control*, pp.438-443, 2003,

[8] Xiaohai Li, Jizong Xiao, Zijun Cai "Backstepping based multiple mobile robots formation control," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1313-1318, 2005

[9] S. G. Tzafestas and G. G. Rigatos, "A simple robust sliding-mode fuzzy logic controller of the diagonal type, " *Journal of Intelligent Robotic Systems*, Vol. 26, pp. 353-388, 1999.

[10] Okyay Kaynak, Kemalettin Erbatur, and Meliksah Ertugrul, "The fusion of computationally intelligent methodologies and sliding-mode control-survey," *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, VOL. 48, pp. 4-17, 2001.

[11] T.-H. S. Li and M.-Y. Shieh, "Switching-type fuzzy sliding mode control of a cart-pole system," *Mechatronics*, Vol. 10, pp. 91-109, 2000.

[12] R. Siegwart and I. R. Nourbakhsh, "Introduction to Autonomous Mobile Robots," *MIT Press*, Cambridge, Mass, USA, 2004.

A. Bazoula, M.S. Djouadi
Military Polytechnic School
P.O. Box 17, 16111 Bordj-El-Bahri, Algiers, Algeria
E-mail: {a,delouahab.bazoula, ms.djouadi}@gmail.com
H. Maaref
IBISC CNRS-FRE 2873
40 Rue de Pelvoux-CE 1455
91020 EVRY CEDEX, FRANCE
E-mail: hichem.maaref@iup.univ-evry.fr

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 185-189

Concern and Business Rule-Oriented Approach to Construction of Domain Ontologies

Crenguţa Mădălina Bogdan, Ana Păduraru

Abstract: A shared ontology is an essential condition for obtaining the stakeholders' agreement as well as for the interoperability of the applications which stakeholders use or are interested in. The paper presents an approach based on a concern and business rule-oriented analysis aimed at partitioning the information system domain in stakeholder-oriented subdomains. A stakeholder's concern depends on his or her interests or preoccupations originating from real world problems. The interests, preoccupations, and concerns depend on the stakeholders' beliefs and knowledge. The concepts in the mental representation of the stakeholders' beliefs and knowledge populate the universe of discourse and make up their vocabulary. The formal description of the intended meaning of this vocabulary forms a domain ontology. The paper exemplifies this approach considering the case of the registration of a new trading company using the services provided by the public administration institutions.
Keywords: information system, business rule, concern, knowledge, ontology

1 Introduction

Fulfilling the interoperability quality attribute is a goal of the developers and users of software and information systems, who have to provide seamless and automatic connections from one system to another, so that systems can work together to solve tasks or problems. Information interoperability spread out on three levels or dimensions: syntactic, structural, and semantic. While for syntactic and structural levels, XML-based standards guarantee interoperability, the semantic interoperability needs semantic models, like ontologies which properly describe the nature of the concepts and their relations.
Designing an ontology is always a challenging task. There are some methodologies [8], which guide the development of ontologies, but their construction is still far from being well-understood. In this paper, we propose an approach based on stakeholders' concerns and business rules of the information systems for designing the ontologies of information systems.

1.1 Information Systems

An information system (IS) is an informational model of one or more work system(-s) of an organization in charge of supplying the information support for the represented system. Therefore, an information system performs the functions of capturing, transmitting, storing, retrieving, manipulating, and supplying data, information, and knowledge [2].

From the kinds of component of the ISs, we focus on business rules. Unfortunately, there is no standard definition of the "business rule" concept. During our research, we used the business rule as a condition that governs the behavior or execution of other elements of the information system: processes, objects, and agents. The condition could be a compulsory constraint that prevents, triggers, and enables the executions of some business processes, or an assertion that defines or constraints some aspects of the organization, or describes some situation and which provides information.

From the business actor's point of view, the business rules state conditions that must be fulfilled if he or she wants to obtain some business service from the organization. For instance, in order to register his or her new trading company at the Trade Register Office (TRO), the founder must provide the records proving the deposit of the stockholders' paid-in capital. Another business rule of the TRO is that any deed of incorporation must be registered by this institution.

In this paper, we propose an approach where given the business rules of an IS we can deduce pieces of knowledge. In addition, in our approach, the pieces of knowledge are derived by an iterative method for every concern. The concern concept and its description are presented in Section 1.2. From the concerns analysis we can identify the beliefs and knowledge of the business actors and workers of the information system considered. Then, the vocabulary of the conceptual domain is created and formally described in a domain ontology that extends a foundational ontology. Throughout our research, we used the top-level ontology DOLCE [7] and one of its modules

D&S [3] which are presented in Section 1.3. The approach is exemplified on a business process of the Romanian public administration (Section 3).

1.2 Separation of Concerns

The separation of concerns is an old decomposing and composing principle that partitions an information or software system into smaller and more manageable and comprehensible parts [5]. Each decomposing criterion is derived from a concern or need belonging to a particular area of interest. Many decomposing criteria are based on stakeholders' concerns.

In [6] we defined the (stakeholder's) concern as a problem-originated care of one or more stakeholders involved in the IS construction or evolution in its natural environment. The care of a stakeholder depends on his/her interest or preoccupation related to a problem from the real world. The interest of a stakeholder frequently results from a need, but can also originate in a desire, a different interest or his/her responsibility in the IS evolution process.

The specification of a concern problem uses a pair of two descriptions: a) the description of the initial state of the current situation, as the stakeholder perceives it, and b) the description of the final state of the situation that reflects the stakeholder's expectations, interests, or preoccupations. Both these two elements are considered as a hypothesis and a conclusion of the problem specification (see example C15).

1.3 Ontologies

According to Guarino's definition, "an ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world" [5]. A conceptualization is a set of conceptual relations defined on a domain space [5]. According to their arity, the conceptual relations may be unary (and they are named concepts) or binary, ternary and these are named relations.

Nowadays, there are some top-level ontologies which describe very general concepts like space, time, matter, object, event, etc., i.e. independent concepts through a particular domain or problem. From these, we used the DOLCE ontology [7] and one of its modules D&S [3].

2 Our Approach

In this paper we propose a concern and business rule-oriented approach to the construction of domain ontologies. The approach has eight steps: 1) the identification of stakeholders; 2) the identification and description of concerns; 3) the concerns analysis; 4) the business rules analysis; 5) the selection of a foundational ontology to be extended by the new ontology; 6) the classification of the concepts according to the foundational ontology; 7) the definition of the ontology; 8) the verification and validation of the ontology. These steps are explained in the subsections below.

2.1 The Identification, Description and Analysis of Concerns

In order to identify the concerns, we start from the analysis of the stakeholders' preoccupations, interests and beliefs, and then we identify how they generate concerns, in other words how the stakeholders think. For this, we resort to contemporary philosophy that provides us with many theories of the mind.

The best known and generally accepted theory is Functionalism which models the states of mind (beliefs, concerns, desires, needs, being in pain, etc.) by considering solely their functional role: transformers of sensory inputs into behavioral outputs, in causal relations with other states of mind [8]. The states of mind are closely related to beliefs and knowledge, which we discuss below.

We consider a belief as a state of mind about a mental representation which symbolizes a mental object depending on a perception [8]. In Cognitive psychology, a mental representation is defined as a psychological mechanism that allows the reflection and the knowledge of an entity, phenomenon, or a state of affairs in its absence. The condition is that this should be previously perceived in the real world [9]. There is a strong relation between knowledge and beliefs: a credible belief accepted by all stakeholders interested in, it is a piece of knowledge.

The identification of knowledge is an analysis activity that focuses on the identification of both the tacit and the explicit knowledge. Furthermore, these pieces of knowledge regard the problem associated with the concern manifested by one or more stakeholders. In our approach, for every concern the beliefs that describe the actual state of affairs of the associated problem must be considered. The description of a part of the actual state of affairs

constitutes the hypothesis of the problem. The questions from the high-level specifications of the concerns also constitute a starting point for the identification of the knowledge and beliefs. However, we do not consider all the beliefs and knowledge of a stakeholder, but only those regarding the explanations of the cause of problems, and related to their concerns.

Each and every concern is described according to the following template: concern code, concern name, high-level specification and roles of the stakeholders, who manifest the concern or are interested in its solution (see example C15).

2.2 The Business Rules Analysis

From each business rule, we can derive one or more pieces of knowledge and/or beliefs. In addition, depending of the type of the IS's component that the business rule constraints, we divided the business rules into six categories: 1) constraints that describe the assertion that a business object must have some property; 2) validation rules for the values or materializations of some property of the business object; 3) constraints that describe the assertion that two or more business objects participate in a relation or a business process; 4) constraints concerning the number of business objects that participate in a relation or a business process; 5) constraints that describe the assertion that a institution or a stakeholder plays some role in some context or business process; and 6) rules imposed by law or state institutions.

2.3 The Process of Writing the Ontology and the Consistency Verification

The mental representations of the stakeholders' knowledge and beliefs are built using concepts that refer to individuals belonging to three categories: physical entities and their relations in the real world, ad hoc conceptualizations based on the stakeholder's experience, and abstract (non-physical or social) entities produced by the human mind and are shared by various communities.

The identification of concepts from every belief and knowledge represents the activity by which a vocabulary is created. The vocabulary is a set of concepts that we use in order to refer to concrete and abstract entities, as well as relations between them from the domains associated with the problems related to the identified concerns. Starting from every belief or knowledge in every concern description, the participating concepts are added to the vocabulary. The vocabulary is used for solving the problem associated with the concern. This activity is repeated until the whole conceptual domain of the problems associated with the concerns shared by stakeholders is obtained.

Then, the foundational ontology is chosen. The foundational ontology can be a top-level see section 1.3) or a domain ontology. Domain ontologies describe a generic domain (like medicine, engineering, history, etc.) specializing the concepts from the top-level ontology adopted.

Subsequently a taxonomy is created by subsuming the concepts to the foundational ontology taxonomy. Then, the domain ontology is created on the basis of the conceptualization of the foundational ontology by formally describing the intension of every concept and its conceptual relations with other concepts. The description is created by imposing restrictions concerning the semantics of the concepts and their relations.

The consistency of ontologies in general is defined as a set of conditions that must be met for the formalization of the model. The conditions assure that none of the definitions of a concept or conceptual relations contradict one another. We consider that the ontology must be verified both by an ontology expert and mechanically, using an ontology editor such as Prot*égé* [4] together with inference tools like RacerPro [11]. This reasoner system allows for the verification the ontology consistency and establishing the risk of logical problems when the ontology is used in a knowledge base to infer new facts. Finally, the ontology should be validated by the users of the information system described by the ontology in order to have the consensus of the stakeholders involved.

3 Case Study

Our approach is applied to the information system of the Romanian Public Administration (short, RPA). The RPA includes public institutions like the TRO, the Public Finance Administration, the Labour Safety and Social Insurances Agency, the Official Gazette Agency, etc..

The RPA also provides services to companies belonging to the business environment or to private entrepreneurs who want to establish their own company. In order to do this, the founder has to register his or her company at the TRO in the city where the company is based. This public institution issues a registration certificate that authorizes the legal operation of his or her company.

In this context, we applied our approach in order to create a domain ontology of the information system of the TRO and the other collaborating public institutions with the aim to provide the Registration certificate of a trading company.

For the first step of the approach, we identified the stakeholders who have a legitimate interest in the information systems considered. They are applicants such as founders, administrators, legal representatives, including companies or corporate entities, and clerks, jurists, judges, operators of service providers such as public institutions and banks.

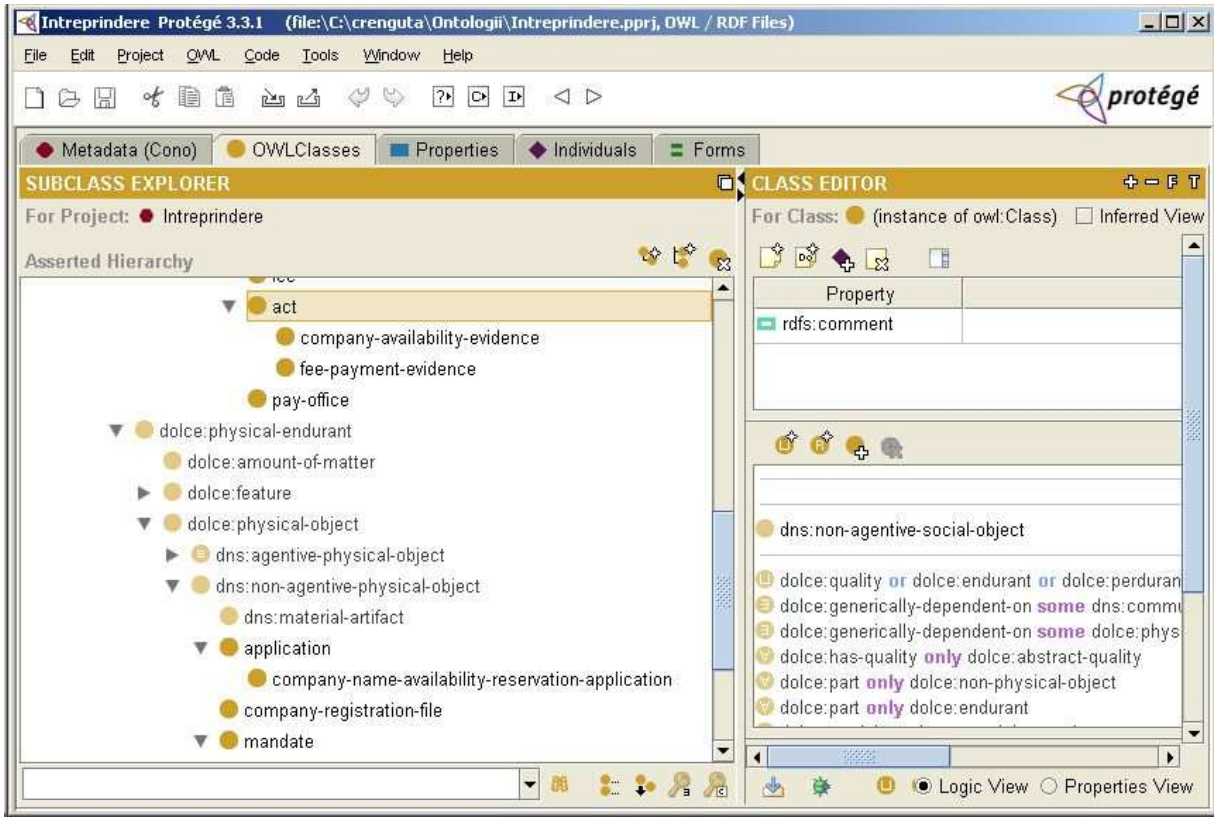


Figure 1: An excerpt of our ontology

For the second step, we identified the concerns of the stakeholders, more precisely 31 concerns of the founder and the other stakeholders. Below, we give the description of a founder's concern.

C15	*Name:* Care to state the new trading company's name	
	Problem	*Hypothesis:* The founder has to choose at least three Romanian names. These names will be verified by the TRO. According to art. 39 Law no. 26/1990 regarding the Trade Register, the names cannot contain certain words. *Conclusion:* What name will the new trading company have?
	Stakeholders: Founder	

The next two steps consist in the analysis of the concerns and business rules. The aim of the analysis is the identification of the pieces of knowledge and beliefs. In the analysis activity, we applied all the heuristics and rules that we provided in the sections 2.1 and 2.2. For instance, in the table below we provide two samples of knowledge and beliefs of the concern C15 .

Code	Description of the mental representation in natural language
B10	Every trading company name may contain the words: "national", "Romanian", "institution" or their derivates subject to the consent of the Government General Secretariat.
K5	All the trading company names are reserved by the TRO.

As a statistical information from the concerns analysis we derived 204 beliefs and pieces of knowledge and from the business rules we obtained 60 beliefs and pieces of knowledge.

Furthermore, from each belief and piece of knowledge we identified the concepts and their conceptual relations. Then, we analyzed them and, using the DOLCE and D&S ontologies, we described the intension of the concepts and their conceptual relations.

In DOLCE, the restrictions are given using a subset of the first-order logic and their verification is a long time task. That is why we translated our domain ontology in OWL DL (Web Ontology Language-Description Logic) language [15] and we checked the ontology consistency with the help of the Protégé tool [4] and the RacerPro reasoner system [11]. In Figure 1 we show an excerpt of our ontology in the OWL language.

4 Conclusions and Acknowledgements

In this paper, we present a concern and business rule-oriented approach for designing an ontology of an information system. This work is funded within the TOMIS project, no: 11-041/2007, by the National Centre of Programs Management, PNCDI-2 - Partnerships program.

References

[1] N. F. Noy, D. McGuinness, "A Guide to building ontologies: Ontology Development 101. A Guide to Creating Your First Ontology", March, 2001 at http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html

[2] S. Alter, "A General, Yet Useful Theory of Information Systems", *Communications of the Association for Information Systems*, vol. 1, 1999.

[3] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, "WonderWeb Deliverable D18. Ontology Library". *IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web*, 2003.

[4] A. Gangemi, P. Mika, "Understanding the Semantic Web through Descriptions and Situations", *Proceedings of the International Conference ODBASE03*, Italy, Springer, 2003.

[5] D. L. Parnas, "On the Criteria to Be Used in Decomposing Systems into Modules", *Communications of the ACM*, 15(12), 1972.

[6] C. Bogdan, L. D. Serbanati, "Toward a Concern-Oriented Analysis Method for Enterprise Information Systems", *Proceedings of the IEEE International Multi-Conference on Computing in the Global Information Technology*, Bucharest, Romania, 2006.

[7] N. Guarino, "Formal Ontology and Information System", *Proceedings of FOIS'98*, Trento, Italy, IOS Press, 1998.

[8] N. Block, ed., *Readings in Philosophy of Psychology*, vol. 1, Cambridge, Harvard, 1980.

[9] M. Zlate, *Psihologia Mecanismelor Cognitive*, Polirom, 2004

[10] Protégé Ontology Editor, http://protege.stanford.edu/.

[11] RacerPro Reasoner, http://www.racer-systems.com/

[12] World Wide Web Consortium. OWL Web Ontology Language Reference. W3C Recommendation, 2004.

Crenguţa Mădălina Bogdan, Ana Păduraru
Ovidius University
Numerical Methods and Computer Science Department
124 Mamaia Blvd., Constanta, Romania
E-mail: cbogdan@univ-ovidius.ro, anna_paduraru@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 190-195

Modeling and Simulation of Short Range 3D Triangulation-Based Laser Scanning System

Theodor Borangiu, Anamaria Dogar, Alexandru Dumitrache

Abstract: In this paper, a simulation environment for a short range 3D laser scanning system that uses triangulation is presented. The simulation is used for integrating a laser scanning probe that uses a line laser and two cameras, with a vertical articulated robotic arm with 6 degrees of freedom and a rotary table. The optical subsystem is simulated using POV-Ray ray tracing software, and the image processing for triangulation, together with the robotic arm kinematics and the user interface, are implemented in MATLAB.
Keywords: 3D Laser Scanning, Simulation, Ray Tracing.

1 Introduction

This work is part of a bigger project, whose goal is to integrate a short range 3D laser scanning probe with a vertical articulated robotic arm with 6 degrees of freedom and a rotary table. The laser probe is able to measure distances from 70 to 250 millimetres, achieving 30 μm accuracy. The robotic arm will move around the workpiece being scanned by using computer-generated adaptive scanning paths, which are computed in real-time while the scanner is discovering the workpiece features. The scanned 3D models will be then reproduced on a CNC milling machine. The simulator presented in this paper is designed to be a development tool and test bench for the adaptive scanning algorithms which will be developed.

2 Motivation

Using a simulation environment for designing the scanning algorithms has several advantages:

- The possibility of collisions between the robotic arm, laser probe and workpiece is eliminated. As the laser probe is an expensive device, collision avoidance is a very important point to consider;
- The system can be analyzed in ideal conditions, with no surface reflections, external light sources or perturbations in the measurements;
- The parameters of the scanning system components, like camera location, focal length, optical sensor resolution, laser beam width, rotary table size and location, can be freely changed, and the influence of these changes can be analyzed thoroughly;
- Scanning of objects with complex shapes and different surface properties, that may not be physically available, is possible.

There are also a number of disadvantages, the biggest one being the computational power involved for accurate simulations in real-time, and the second one being the difficulty for simulating the less-than-ideal conditions of the real system. In the authors' opinion, the advantages outweigh the shortcomings involved here.

3 Problem description

The laser scanning system that is to be simulated consists of three main components:

- The robotic arm
- The laser probe
- The rotary table

For the robotic arm, the simulator should allow displaying the robot in any user-defined position. The user should be able to control either the joint angles for each articulation, or the Cartesian position together with the orientation given by ZYZ' Euler angles. The user should also be able to specify the tool transformation.

For the laser probe, the software should simulate the interaction of the laser beam with any user-defined workpiece, having various surface properties. The two cameras which are integrated into the laser probe should also be simulated, and the image which would be captured by them should be displayed to the user. Furthermore, the laser beam should be detected in the images from the two cameras, and the triangulation equations should be applied to them in order to compute a point cloud. The point clouds obtained from simulating the scanning process from different viewing angles should be transformed into a fixed coordinate system, that will be attached to the workpiece being scanned.

As for the rotary table, its rotation have to be simulated, and the workpiece that sits on the table should rotate synchronously with the table. Behaviors such as inertial movement due to fast table movements do not have to be simulated; the workpiece should be considered attached to the table.

4 Proposed Solution

The first decision is to choose which software environments are suitable for development of this simulator. Since it is needed to model the interaction between the laser beam and an arbitrary surface, a possible solution is to compute the intersections between the laser rays and the given object. This is not a trivial task, and there are many 3D renderers that use the *ray tracing* method, for which there is a good description in [3]. Examples of ray tracing software include the free POV-Ray, Rayshade, Radiance, and the commercial software, such as 3D Studio Max, Maya or Catia.

For this simulator, POV-Ray was chosen for modeling the laser beam. POV-Ray uses a *Scene Description Language* [4], which is used to describe the objects, lights and cameras that interact in a virtual environment by means of ASCII-based input files. This capability allows easy interfacing with many programming environments. Furthermore, there is a command-line version of POV-Ray, that simplifies integration of POV-Ray with other software programs.

For the computational side of the simulator, which involves 3D matrix multiplications and trigonometric operations, the programming environment was chosen to be Matlab. Its advantages include an interpreted language with syntax oriented for matrix operations and mathematical functions, also having an Image Processing Toolbox and built-in 2D/3D graphics capabilities.

4.1 Robot arm and rotary table modeling

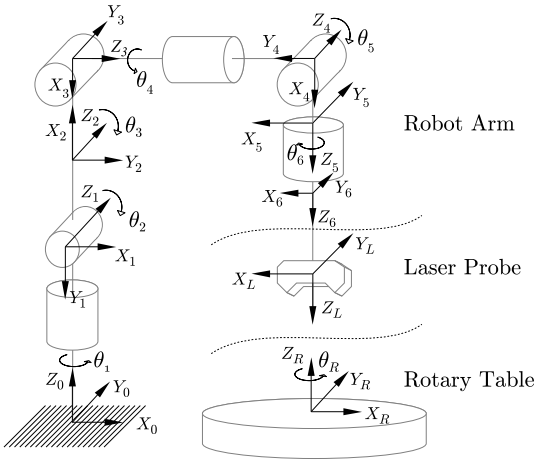
For modeling the robot arm, there are two main points to consider: computing positions and orientations for every link of the robot, including direct and inverse kinematics, and displaying the robot to the user in a given configuration.

The World coordinate system used here, $X_0Y_0Z_0$, is right-handed, with the X_0Y_0 plane being horizontal, X_0 axis pointing forward, Z_0 axis pointing upwards, and the origin being at the base of the robotic arm. The rotary table has the same reference frame as the workpiece, $X_RY_RZ_R$, and the Z_R axis points in the same direction as Z_0.

Direct and Inverse Kinematics

The direct kinematics for the robot arm function is obtained by using the Denavit-Hartenberg [1] convention. The first step is to assign individual reference frames to each link from the kinematic chain, which includes the six robot arms and the laser probe (Fig. 1(a)). The direct kinematics function is the product of the individual homogeneous transformations [2] for each link $i = \overline{1..6}$. An individual matrix, called T_i^{i-1}, is the transformation from the $(i-1)^{th}$ link reference frame to the i^{th} link reference frame. The 0^{th} link is the robot base, and the 7^{th} link is the laser probe. T_L^6 is the transformation from the 6^{th} joint (robot flange) to the laser probe reference frame, and it is a constant matrix, since the laser probe is rigidly attached to the robot. Knowing the the Denavit-Hartenberg parameters a_i, d_i, α_i and θ_i for each joint $i = \overline{1..6}$, with θ_i being the joint variables, the individual transformations T_i^{i-1} can be written:

$$T_i^{i-1} = \mathcal{R}_Z(\theta_i)\,\mathcal{T}(a_i, 0, d_i)\,\mathcal{R}_X(\alpha_i) \tag{1}$$



(a) Reference frames for robot arm links, laser probe and rotary table

Link	a_i [mm]	d_i [mm]	α_i [deg]	θ_i [deg]
Rot	–500	–200	0	$-\theta_R$
1	75	335	–90	θ_1
2	270	0	0	θ_2
3	–90	0	90	θ_3
4	0	295	–90	θ_4
5	0	0	90	θ_5
6	0	80	0	θ_6

(b) Denavit-Hartenberg parameter table

Figure 1: Reference frame assignment and Denavit-Hartenberg parameter table

where $\mathcal{T}(x,y,z)$ is the homogeneous translation, and $\mathcal{R}_A(\phi)$ is the homogeneous rotation around axis A with angle ϕ. All the distances are expressed in millimetres, and all the angles are expressed in radians.

The direct kinematics transforms are 4×4 matrices:

$$T_{DK}^{*} = T_6^0 = T_1^0\, T_2^1\, T_3^2\, T_4^3\, T_5^4\, T_6^5 \tag{2}$$

$$T_{DK} = T_7^0 = T_{DK}^{*}\, T_L^6 \tag{3}$$

The matrix T_{DK}^{*} may be used for transforming any object attached to the robot flange, in this case, the laser probe. The matrix T_{DK} expresses the position of the laser reference frame, with respect to the robot base, and will be used to convert the laser probe measurements into an unique coordinate system.

The inverse kinematics was implemented using Peter Corke's Robotic Toolbox, function `ikine`, which uses the pseudo-inverse of jacobian method [6]. The simulator is also able to connect to an existing robot controller via TCP/IP and use its internal inverse kinematics routine, which has been found to have a slightly different behavior, depending on the initial estimation.

Considering the rotary table as a 7-th link of the kinematic chain, and moving the reference frame to the center of the table $X_R Y_R Z_R$, the table is modelled as a new link, applied before the 6 links of the robot (Fig. 1(b)), and the transformations required to convert from the robot base reference frame to the rotary table reference frame and viceversa are:

$$T_0^R = \mathcal{R}_Z(-\theta_R)\, \mathcal{T}(a_R, 0, d_R) = \mathcal{R}_Z(-\theta_R)\, \mathcal{T}(-500, 0, -200) \tag{4}$$

$$T_R^0 = \left(T_0^R\right)^{-1} = \mathcal{T}(-a_R, 0, -d_R)\, \mathcal{R}_Z(\theta_R) = \mathcal{T}(500, 0, 200)\, \mathcal{R}_Z(\theta_R) \tag{5}$$

Rendering

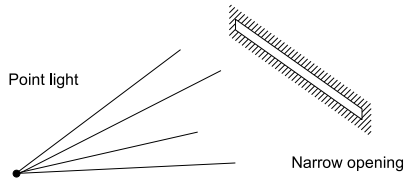
The robot is displayed by rendering it using POV-Ray, from a triangle mesh model obtained from the CAD model for the robot arm. For the robot arm used in this project, the CAD model is available for download from the manufacturer's web site [5]. Each link of the robotic arm is modeled as a triangle mesh, and can be freely moved in the scene. Using the joint values computed from Inverse Kinematics routine, the links are placed in the correct position using the individual transformations T_i^{i-1} from Eq. 1, and the scene is rendered, with the result being shown in Fig. 4(a) from Section 5. A similar approach was used for rendering the laser probe and the rotary table, by creating POV-Ray models and applying the transformations T_{DK}^{*} and T_R^0 from Eq. 2 and Eq. 4.

4.2 Laser probe modeling

The reference frame of the laser probe is $X_L Y_L Z_L$, from Fig. 1(a), and it will be called *XYZ* in this section in order to simplify the notations.

Laser beam modeling

The laser probe emits a laser beam focused into a plane, such as when it intersects a surface, it casts a line or a curve. This laser beam may be modelled as a point light source, which is constrained to pass to a narrow opening (Fig. 2). In the example from Fig. 2(b), the laser rays start from origin, is projected in the positive direction of the Z axis, the plane of the laser rays is YZ and the narrow opening has the dimensions L (large edge) and W (small edge) and is located at a distance D from the origin.



(a) Laser beam modeling principle

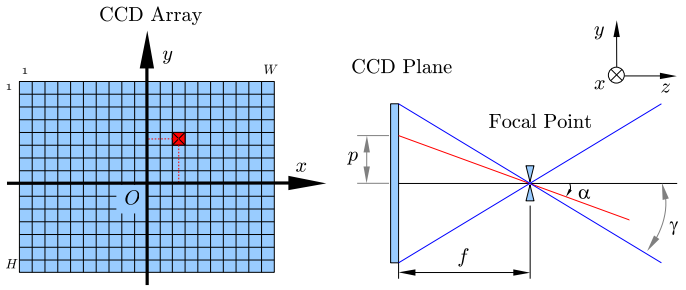
```
light_source {
    <0, 0, 0>
    color rgb <1, 0, 0>
    projected_through {
        box {
            <-W/2, -L/2, D>,
            < W/2,  L/2, D + eps>
        }
    }
}
```

(b) POV-Ray example

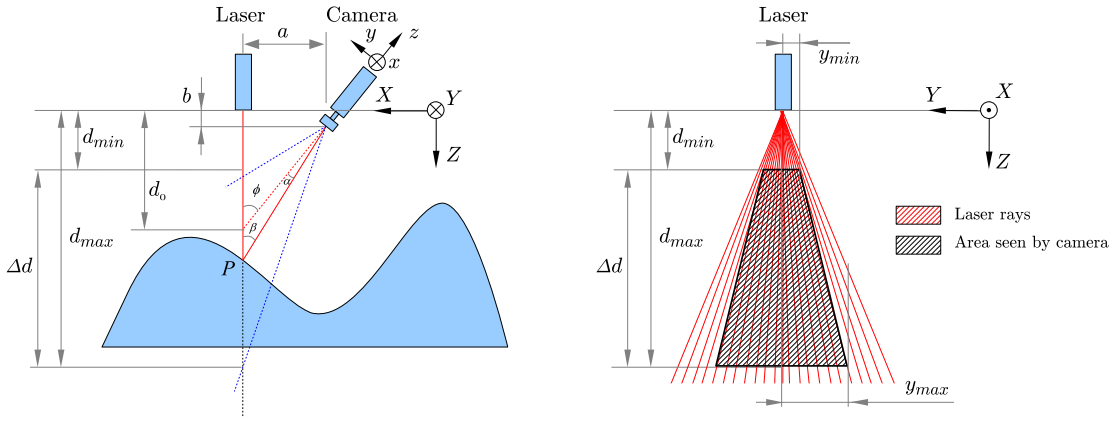
Figure 2: Laser beam modeling using POV-Ray

Camera modeling and triangulation

The cameras used in the laser probe are be modeled as two standard perspective cameras, which may be implemented in POV-Ray by entering their parameters such as position, orientation and focal length. In the following text, only one of the two cameras will be described, as the other one is identical and symmetrical to the first one.



(a) CCD sensor and its reference frame



(b) Side view of the laser probe

(c) Front view of the laser probe

Figure 3: Triangulation

Let XYZ be the reference frame of the laser probe (Fig. 3(b) and (c)), and let xyz be the reference frame of the CCD array from the camera (Fig. 3(a) and 3(b)). Referring to Fig. 3(b), the camera position and orientation with

respect to the laser device is given by the parameters a, b and ϕ.

Using these notations, let $P = (P_X; P_Y; P_Z)$ the point of reflection of a laser ray, in the *XYZ* reference frame, and let $p = (p_x; p_y)$ be the coordinate of the pixel at which the ray was detected on the CCD matrix, in *xy* reference frame. Knowing the pixel coordinates p, the location of the point P can be expressed using the triangulation equations (6):

$$P_X = 0 \qquad P_Y = \frac{a}{f\ \sin\left(\phi - \arctan\frac{p_y}{f}\right)}\ p_x \qquad P_Z = \frac{a}{f\ \tan\left(\phi - \arctan\frac{p_y}{f}\right)} + b \tag{6}$$

where $f = \dfrac{H}{2\tan\gamma}$ if the unit length is considered to be 1 pixel, i.e. the distance between two adjacent pixels on the CCD array.

The area in the plane determined by the laser rays, i.e. *YZ* plane in Fig. 3(c), is a trapezoid, and its limits are z_{min}, z_{max}, y_{min} and y_{max}, whose expressions are given in Eq. 7. The scanning range is thus given by z_{min} and z_{max}, and the length of the laser line L_L that is effectively being analyzed is dependent of the distance of the scanned object with respect to the laser probe, and varies from $L_{Lmin} = 2\,y_{min}$ when the workpiece is close to the probe, to $L_{Lmax} = 2\,y_{max}$ when the workpiece is far from the laser probe.

$$y_{min} = \frac{a\ \tan\gamma}{\sin(\phi+\gamma)} \qquad z_{min} = \frac{a}{\tan(\phi+\gamma)} + b \tag{7}$$
$$y_{max} = \frac{a\ \tan\gamma}{\sin(\phi-\gamma)} \qquad z_{max} = \frac{a}{\tan(\phi-\gamma)} + b$$

Creating the point cloud

There are three main steps in computing a point cloud by processing the image obtained from the simulated camera. In the first place, one needs to identify the laser beam in the image. The simplest and fastest method is to use a gray workpiece and a red laser, and apply a thresholding operation on the saturation component of the image. The second step is to apply Eq. 6 to every pixel on the laser beam, obtaining a cloud point in the laser probe's local reference frame. The third step is to transform the coordinates of the points in an reference frame that is attached to the workpiece. The homogeneous transformation required here is computed by first expressing the point cloud coordinates in robot's World reference frame, and then expressing them in the rotary table's reference frame, using the matrices defined in Eq. 3 and Eq. 4:

$$T_{PC} = T_R^0\ T_{DK} \tag{8}$$

A more advanced image processing strategy would be able to achieve sub-pixel accuracy, and may also address the reflections, ambient light and other error sources.

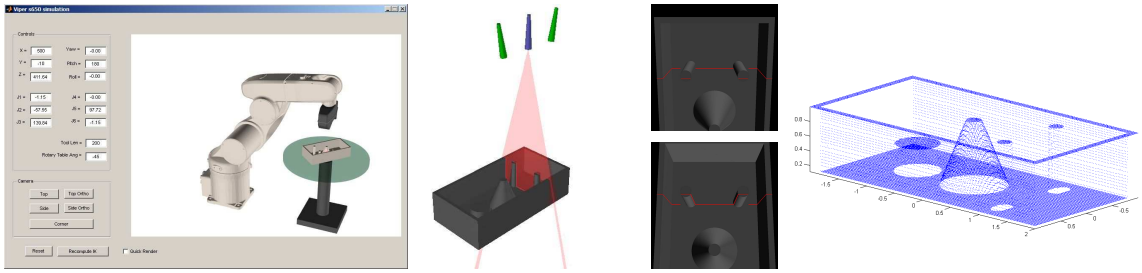
5 Simulation example

The screen shot of the simulation software is displayed in Fig. 4(a). The user can control either the position and orientation of the robot in Cartesian mode, relative to either robot base or rotary table, or the individual joint angles. The software may simulate a continuous movement of the laser probe over the workpiece, compute the images that would be seen by the two cameras, and generate a point cloud from processing them (Fig. 4(b)-(d)).

6 Summary and Conclusions

A computer simulator of a 3D laser scanning system was presented, along with the theoretical aspects involved. The software simulates the kinematics of the robot arm moving synchronously with a rotary table, and the interaction of the laser probe, which uses a laser beam and two cameras, with a virtual workpiece. A point cloud representing the workpiece can be generated by processing the rendered images. Other types of robot arms may be simulated by providing their Denavit-Hartenberg parameters and a 3D mesh model for displaying.

The bottleneck in the simulation is the rendering speed, which is currently performed by ray tracing. Therefore, real-time simulations are not currently possible. On an Intel Pentium 4-M processor at 2.0 GHz, the simulator is able to render and process around 3 frames per second by simulating a 160×160 sensor, while the speed of a real laser probe may be of the order of 100 frames per second, as described in the technical specifications.



(a) Screen shot of the simulator (b) Schematic view of laser probe (c) Images from the two cameras (d) Computed point cloud

Figure 4: Laser probe simulator

References

[1] Mark W. Spong et.al., *Robot Modeling and Control*, John Wiley and Sons, Inc., pp. 71-83, 2005

[2] Tom Davis, *Homogeneous coordinates and computer graphics*, 2001

[3] Andrew S. Glassner, *An Introduction to Ray Tracing*, Morgan Kaufmann Publishers, Inc., 1989.

[4] Persistence of Vision Raystracer Pty. Ltd., *POV-Ray Online Documentation*

[5] Adept Technology, Inc. Web Site, *www.adept.com*

[6] P.I. Corke, A robotics toolbox for MATLAB. *IEEE Robotics and Automation Magazine*, Vol. 3, No. 1, pp. 24-32, March 1996.

Theodor Borangiu, Anamaria Dogar, Alexandru Dumitrache
University Politehnica of Bucharest
Department of Automatic Control and Industrial Informatics
Splaiul Independentei 313, Bucharest, Romania
E-mail: borangiu@cimr.pub.ro, dogar@cimr.pub.ro, alex@cimr.pub.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 196-200

Device for Protection to the Lack of the Pulse for the Tri-Phase Rectifiers in Bridge

Ilie Borcosi, Onisifor Olaru, Nicolae Antonie

Abstract: In this work is proposed a device for protection at the lack of pulse for the tri-phase rectifiers in bridge with load inductive resistive (*dc motors*).
Keywords: protection, device, no pulse, tri-phase rectifiers.

1 Introduction

The device is used in case of action system with the DC motors for the protection of the static converters and motor. This device makes the object of a demand of license of invention. The invention refers to a method and a device for the protection of rectifiers when the semiconductor element is not working and overload. When a semiconductor element is not working they can be shown because of the defection of the command circuit (*does not receive the command signal*) or because of his defection. The idea of making this kind of device has appeared as a necessity for the predictive maintenance of action system with the DC motors.

There are known diverse circuit for the protection of rectifiers: at the lack of phase, at overload, over voltage etc. The most recent follow the voltage between the force terminals of electronic devices, interfering when this fall of tension is the bigger then the one of saturation, protecting at the exceeding of normal current. These circuited present the disadvantages that they don't detect when the semiconductor device is not working, when he must be in conduction (*when receives the command*).

The purpose of the device is to determine when the semiconductor devices of rectifiers are not working to the fellowship to the form of current through load and the lack of pulse.

2 Presenting the device

The figure of the protection device for a tri-phased rectifier in bridge is presented in the figure 1 where you can see clearly the following components:

- CT1 and CT2 -current transformers (current transducer) for the measurement of current for 2 phases of the rectifies;
- RP1, RP2, RP3-rectifiers of precision made from operation amplificatory;
- $\overline{\sum}$-summator change-over switch for the remaking of the current of the 3 phase for the sum of the received signals for precision rectifiers;
- decision block.

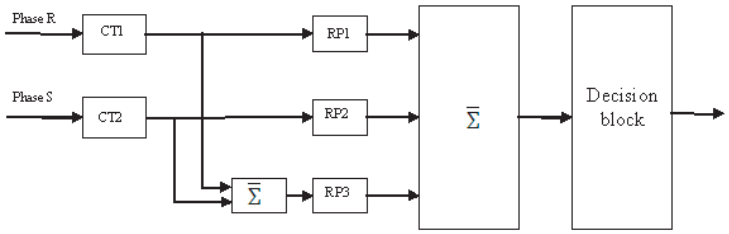


Figure 1: Block scheme

At the exit of the redress circuit and sum, made with AO [1] is obtained the signal u_{sum}, which has the same form of variation with the current through the DC motors.

The decision bloc can be made using the analogical circuit, logical circuit or both. It must work the signal received from the summator, in the purpose of verifying the lack of pulse and/or verifying the rise above the admissible current through the load.

3 The making of the decision bloc using the analogical and logical circuit

The bloc contains: the filtration circuit , comparing , timing ,counting the lack of pulse and a exit floor. The filtration circuit doesn't leave the current component being made with the operational amplification. The inferior limit of the frequency of the amplification band is determent by the passive elements from the entrance circuit are obtained from the relation for pulsation:

$\omega_j = 1/(RC) \Longrightarrow 2\pi f_j = 1/(RC) \Longrightarrow f_j = 1/(2\pi RC)$

Being only six pulse in a period of 20 ms (for the frequency of 50 Hz) it must be used the condition that the inferior limit of frequency of the filtration circuit is smaller then the frequency of the pulse: $fu_{sum} = 6 * 50Hz = 300Hz$

Choosing $f_j = 200$ Hz and $C = 0,1\mu F$ will result the standard value for $R = 8.2k\Omega$.

The filtration circuit also has the comparison function; the operational amplifier is working in an open loop, without reaction, so that the signal from the exit of the comparator will change at every passing through zero of the signal from the exit of the filter.

The signal from the exit of the filter, u_f, together with the signal u_{sum} and the signal u_c from the exit of the comparator are exposed in the figure 2, where no pulses are missing, also in the figure 3, where the pulses are missing .

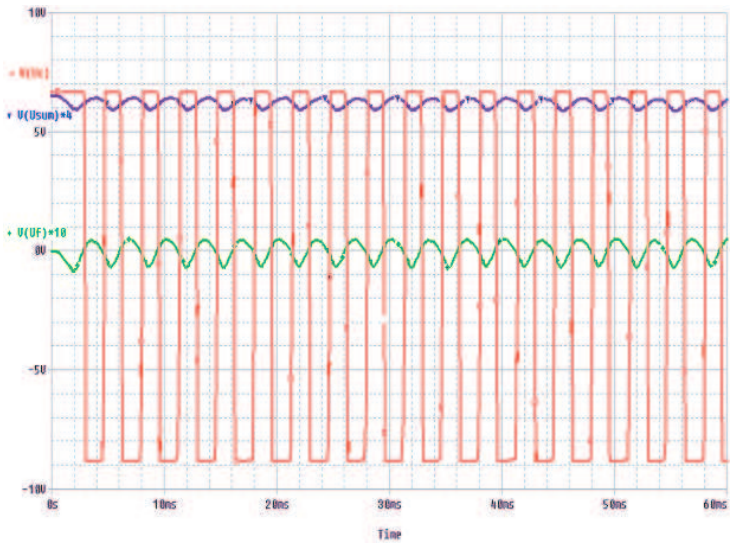


Figure 2: The signal u_f, u_c and u_{sum} when the pulses are not missing

The signal from the exit of the filter uf transforms into rectangular impulses with the help of a comparator, after that he rectifies after which it applies the timing circuit made with the retriggerable monostable circuit.

The monostable period is chosen at 5 ms, so that these being bigger then the duration of the pulse (20ms/6 pulses=3,3ms) and smaller than the duration of the two pulses missing (2*3,3ms=6,6ms).

Because the monostable is retriggerable configured according to those shown earlier the exit signal from the monostable if no pulses are missing, they should be at level one logical. Then when no pulses are missing these signals will change between one and zero logical.

The exit signal from the monostable um is presented in the figures 4 and 5, togheter with the signals uf, the exit signal from the rectifier ur, and the signal usum, for the situation when pulses are not missing, and when the pulses are missing.

A electronic device that doesn't work determents the missing pulse until this situation that doesn't work is eliminated.

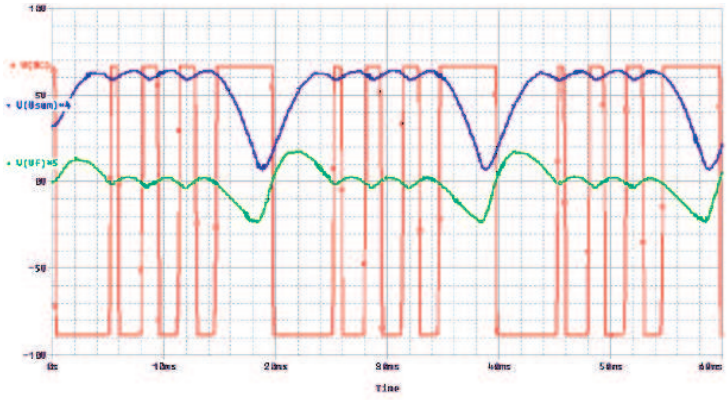


Figure 3: The signal u_f, u_c and u_{sum} when the pulses are missing

There is a possibility to appear the situation when the sporadic pulses are missing (*for example the sudden modification of the prescribed tension at the commanded rectifiers*), this is why we used a counter which will count the missing pulse in a interval of time after which it will reset.

If we consider the load of the rectifier is a direct current motors and the speed changes suddenly , then the signals from the out of the filter , uf, can have the shape in figure 6, which coincide with situation for missing pulses for monostable.

The elimination of these situation is made using a binary counter, which will count the positive transition of the impulses from the Q out of the monostable. If the pulses are missing at the Q out of the monostable it will obtain a impulse in 20 ms (*one transition from 0 to 1*).

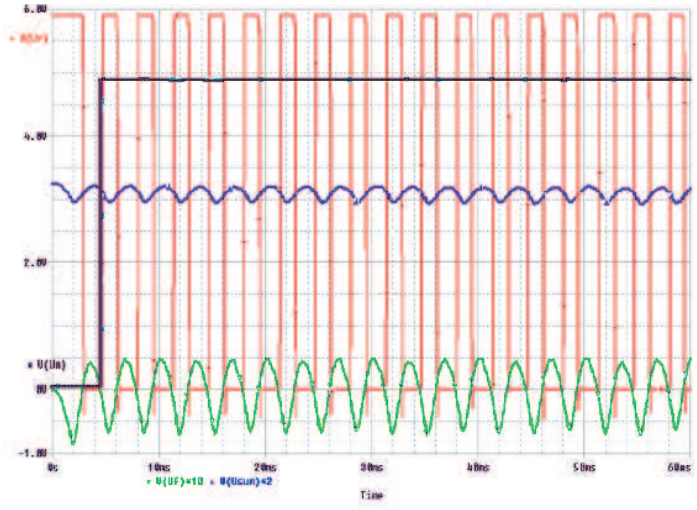


Figure 4: The signals u_f, u_r, u_m and u_{sum} when pulses are not missing

We taken in consideration all 4 outputs of the counter, which are tied to the entrance of a AND gate. If the pulses are missing at the exit of the gate it will be obtained 1 logical after 15 impulses applied at the entrance of the counter which means after 15*20ms=300ms. If we want a smaller interval of 300 ms, we connect at the entrance of a AND gate, the outputs counter which mace up to the wanted number. If we want a bigger interval (*for a load of bigger power*), a secondary counter is connected one after an other (*integrated circuit CD4520 has two counters*).

So that we don't count all the isolated situation of no pulse, we made a reset circuit of the counter, which brings the outputs of the counter to zero, if no pulse is missing in a interval of time equal to 40 ms.

The reset entrance of the counter is active on 1 logical. The reset command circuit is made with a retriggerable

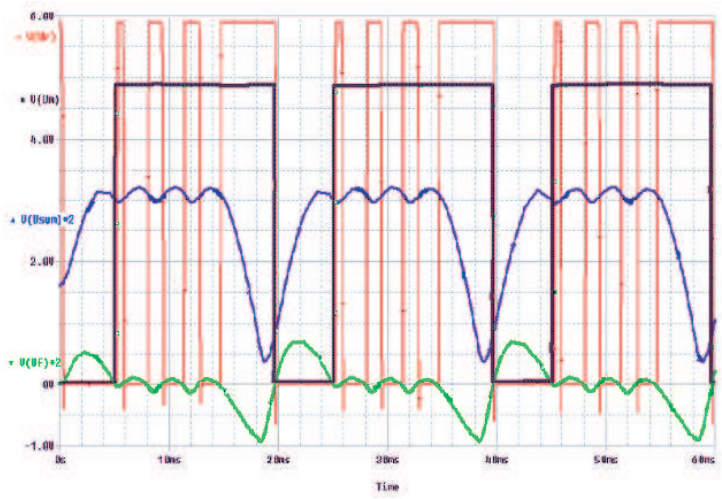


Figure 5: The signal u_f, u_r, u_m and u_{sum} when the pulses are missing

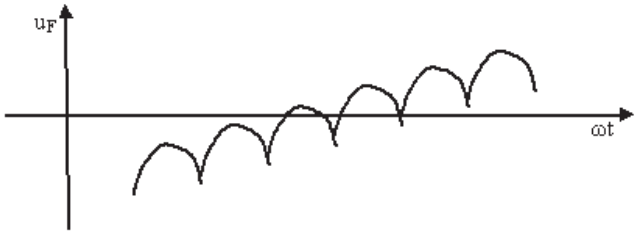


Figure 6: The signal uf when the current through the load suddenly changes

monostable, with the impulse width 40 ms (*two period of 20 ms*). The out impulse of the reset monostable is delayed with the help of the RC circuit. The condenser time loading constant is t1=R*C=40ms, and the time constant of overload is approximately 1 ms. The counter also resets at the coupling of the feeding tension with the help of a RC circuit.

So, at the out of the counter, better said at the exit of the and gate, we obtain 0 logical when no pulses are missing. When the pulses are missing the exit of the and gate will change in 1 logical. These situation can be memorized with the help of a logical circuit so that the operator after realizes that the pulses are missing through the lighting of a led to resolve the problem and reset the circuit by pressing the reset button.

For the projecting of the memorizing circuit we applied the synthesis procedure of the automat asynchronous theory according to the Huffman method and we obtained the circuit from the figure 7, where x1 ties to the out of the counter, and Z is the out of the circuit.

To detect the exceeding of nominal value of the current we use two operating amplifiers: one as a reversor and the other as a comparator. The medium value of the current applied on the inverting entrance of the first operator and on the non inverting entrance of the second one we applied a value of the overload current which we detect with the help of a potentiometer.

When the current rises above the overload value, the out of the operational amplifiers commutes in a bigger positive potential which will polarized directly the diode from the positive reaction.

In these case the diode enters in conduction, being directly polarized bringing a part of the positive potential from the out of the inverting entrance, which forces the exit to stay at a big positive potential (*the memoring of the defect*).

We use only one reset button with two normal opened contacts: one for the detection of the missing pulse circuit and the other for the circuit of overload.

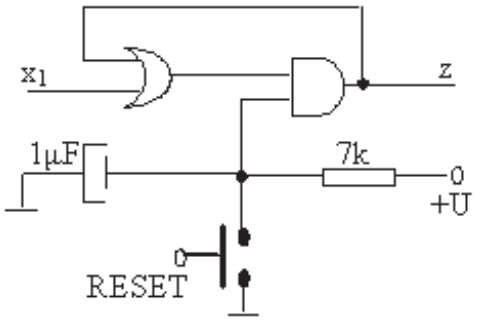


Figure 7: The structure of the memorizing circuit

4 Conclusions

The two defection condition can be seen with the help of the led and then they are put in an OR gate. To increase the exit current of the device we connect on exit floor (a transistor).

The application and the using of the device has the following advantages:

- is basically easy;
- it eliminates the protection circuit and no phase and overload;
- it ensures the protection at no pulse (*the semi conductor element of the redresser not working*);
- it can be used for wanted rectifiers and but also for unwanted;
- also, it can use any type of mono or try phased rectifier, with the exception of the monophased monoalternance rectifier, modifying only the monostables duration;
- it can also be used as a current transducer, giving the information about the medium value of the current trough the load;
- the biggest advantage is represented in the case when the rectifier is a part of a automat relation system of revolutions, of a DC motors. In this situation if the pulses are missing, by the diminution of the medium value of the exit tension from the rectifier ant in the same time the charge tension of DC motors and revolutions. But by not being modified the revolutions is maintenance constant, by increasing the remaining pulses (*the medium value is being kept constant*) which leads to the increase of the current amplitude above the admissible value (*especially if 4 pulse are missing*).The actual circuits don't make this protection when there is no pulse.

References

[1] Olaru O., *Amplificatoare integrate in echipamente de automatizare*, Ed. Universitaria Craiova, 2003.

[2] Manolescu A. M., Manolescu A. Mihut I., Muresan T., *Circuite integrate liniare*, Ed. Did. si Ped. Bucuresti , 1983.

[3] CD4520-CMOS Dual Up-Counters, Texas Instruments, *www.ti.com*.

Ilie Borcosi, Onisifor Olaru, Nicolae Antonie
Constantin Brancusi University
Department of Automatic and Informatics
Bld. Republicii, Nr. 1, 210152 Târgu-Jiu Romania
E-mail: {ilie_b,olaru,nicolae.antonie}@utgjiu.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 201-205

A New Method for Macroflows Delimitation from a Receiver's Perspective

Darius Bufnea

Abstract: This paper presents a new approach for shared bottlenecks detection from a receiver's perspective. This approach uses flow clustering at the receiver, based on passive observations of inter-packet arrival time intervals. We also suggest a new cost function useful in the flows clusterization process into macroflows. The proposed method can be used in the discovery of path patterns or for extending the macroflow granularity in an improved Congestion Manager.
Keywords: congestion control, bottleneck, Congestion Manager, macroflow.

1 Introduction

The latest years have brought changes in the dominant traffic type carried by the Internet infrastructure. In the mid-80s the dominant traffic was generated mainly by specialized "well-behaved" users located in universities or in research centers. Later, in the 90s, the source of the Internet traffic migrated towards business and residential users. Although, the profile of the Internet user has not changed in the last decade, the traffic profile and the amount of traffic increased dramatically. This was mainly caused by new Internet applications such as: multimedia streaming applications (video on demand over Internet, radio broadcasts, voice over IP) and peer-to-peer applications, used mainly for exchanging huge amount of data (e.g. multimedia files of considerable sizes). The traffic generated by the latest Internet applications, used mainly by unspecialized users, is carried sometimes over non congestion-aware protocols such as UDP. Consequently, the network and the transport layers must make continuous efforts to keep fairness among all Internet users, keeping their "best-effort" traffic in normal throughput patterns, while performing smooth congestion avoidance.

Congestion avoidance and control in Internet is done either at network level inside transport routers [1] or at protocol level inside a peer's TCP/IP stack [2]. It is desirable for each transport protocol to implement a congestion avoidance algorithm, or if such an algorithm is not available in the transport layer (for example the UDP transport protocol lacks a congestion avoidance algorithm), it must be implemented in the higher application layer. The congestion control at the network level is required to interfere when a peer in the communication process is not congestion-aware and the amount of data it injects into the network exceeds the amount of data carried by the network for concurrent streams.

2 Previous Work

The current congestion control mechanism, as specified and implemented by the TCP/IP stack [3] performs per-flow congestion checking. For each active TCP/IP connection, the network stack computes individually and maintains separately a series of state variables such as: congestion windows size, round trip time or retransmission time-out. The maintenance of these variables is done on a per-flow basis - there is no coordinated management of streams which compete with each other for scarce bandwidth, rather then sharing the state variables' values whose computation is often redundant. There is a proposed mechanism that implies collaboration between streams that share the same congestion parameters and often, in this situation, the same state variables values in a so called macroflow. Such a collaboration between streams would be managed by a Congestion Manger [4]. However, the problem remains in identifying the flows that must face the same congestion situation. One set of such flows are those that share the same network bottleneck.

There are some proposed techniques that detect shared bottleneck links by using additional injected traffic in the network ([5], [6]). Such an approach has the disadvantage that increases network load and depends on features that might not be available everywhere. Other techniques presented in [7] and [8] detect common bottlenecks by inspecting the time evolution of state variables and clustering flows based on this evolution. However, these techniques can be used only from the sender's point of view. In [9] the authors suggest an information based theory solution that can be used from the receiver's point of view.

In this paper, we suggest a mechanism similar to the one presented in [9] for identifying from the receiver's perspective the set of flows that share the same bottlenecks in their path towards the same destination but use a different technique for flow clustering and a different distance (similarity) measure. Such a set of flows is suitable for extending the granularity of constructing a macroflow in an improved Congestion Manager [4].

3 Mechanism and Data Model

The approach presented in this paper is suitable for detecting shared bottlenecks from a data receiver's point of view or from any transit router's point of view. The detection of shared bottlenecks is done passively by simple observation of flows' inter packet arrival time. Data packets generated by two or more sources that transmit data to a receiver over the same bottleneck, will arrive to destination at equidistant moments in time. This is because the bottleneck router has its queue full and it injects periodically, at constant length time intervals, packets in the congested output line. But these packets that transit a common bottleneck will mix up with data generated by other hosts. So, the main problem is to identify those sources that generated packets which arrive to receiver at equidistant moments in time.

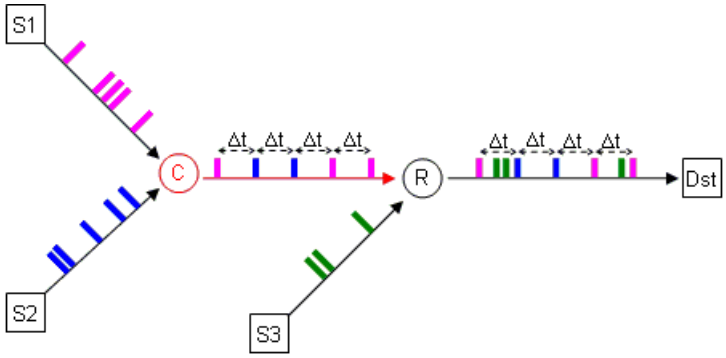


Figure 1: Equidistant inter-packet arrival time intervals

For a better understanding of the mechanism above we analyze it for the network depicted in figure 1. The S_1 and S_2 sources send packets to the *Dst* receiver over the same congested link *C*-*R*. The congested router *C* queues incoming packets and sends them over the congested link at a constant rate, each packet is output in a Δt time interval. However, this regulated traffic is mixed on the *R*-*Dst* link with the traffic generated by the S_3 source. So, at the *Dst* receiver, we have to identify those packets that reach the destination at equidistant moments in time, i.e. packets generated by the $\{S_1, S_2\}$ source set.

The data model used in this paper was presented in [9]. However, we use a modified version of the data model, adapted to achieve our goals. Let $D = \{d_1, d_2, \cdots, d_n\}$ be the set of incoming data packets at the *Dst* receiver. This set is ordered by the incoming timestamp of each packet $t_{d_1} < t_{d_2} < \cdots < t_{d_n}$ where t_{d_i} is the timestamp when *Dst* receives the d_i packet. The D packet set is generated by the source set $S = \{S_1, S_2, \cdots, S_m\}$, usually $m << n$. This assumption is obvious and respects the real traffic patterns. We denote by $Source(d_i) \in S$ the source host of packet d_i and by $\Delta t_{de} = |t_d - t_e|$, $d, e \in D$, the time interval spent between the arrival time of packet d and the arrival time of packet e. Let also F_i be the flow generated by source S_i, $T_i = \{t_{i_1}, t_{i_2}, \cdots\}$ the arrival times of flow F_i's packets and $\Delta_i = \{\delta_{i_1}, \delta_{i_2}, \cdots\}$ the inter-packet arrival time intervals of flow F_i, $\delta_{i_1} = t_{i_2} - t_{i_1}, \delta_{i_2} = t_{i_3} - t_{i_2} \cdots$. Each of the values above can be easily identified by a TCP/IP receiver by direct observation of incoming packets.

In order to simplify our model we made the assumption that all incoming packets at the receiver have the same size. For packets of different sizes, the output intervals can be normalized by the capacity of the congested link. For instance, if two different packets d and e have different sizes, B_d and B_e, those two packets will be output over the congested link at $B_d/C \approx B_e/C$ time intervals, where C is the capacity of the output link.

Identifying those packets that share the same bottleneck means determining at least a subset $D' \subset D$ of packets, $D' = \{d_{i_1}, d_{i_2}, \ldots, d_{i_k}\}$, $k < n$, ordered by arrival timestamps ($t_{d_1} < t_{d_2} < \ldots < t_{d_n}$), so that $\Delta t_{d_{i_1} d_{i_2}} \approx \Delta t_{d_{i_2} d_{i_3}} \approx \ldots \approx \Delta t_{d_{i_{k-1}} d_{i_k}}$. Also, for each such set $D' \subset D$, we can also identify $S' \subset S$ subset of sources , $S' = \bigcup_{d_{i_k} \in D'} \{Source(d_{i_k})\}$. All the flows generated by the sources in S', share the same bottleneck and will be considered part of the same macroflow.

4 Algorithm

Our main goal is to group in a cluster those flows that share the same bottleneck to the destination. Without considering the mixed data traffic, the bottleneck is reflected in a constant inter-packet arrival time intervals at the receiver. Each flow in the clustering process will be represented by its inter-packet arrival time interval vector (Δ_i), while the correct cluster will be represented by a merged vector of inter-packet arrival time intervals. We will use for clustering a hierarchical algorithm that uses as the cost function the standard deviation of a vector. For a cluster C, with a merged data vector Δ having K components, its cost is represented by:

$$\sigma_C = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\Delta_i - \bar{\Delta})^2} \tag{1}$$

In figure 2 we present the image of a correct determined cluster vs. an incorrect one.

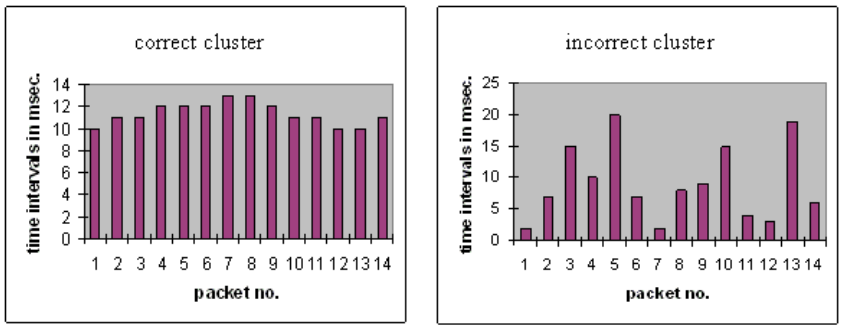


Figure 2: A correct identified macroflow vs. an incorrect one

We are using an iterative clustering algorithm, from the K-means family. The algorithm tries to group packets in clusters using the approach above, in order to reduce clusters final costs.

```
Subalgorithm ClusterIdentification is:
Input:
   m - the total number of flows that reach us;
   F_i, i = 1..m - flow i;
   T_i = {t_i1, t_i2, ...,} - the arrival times of flow F_i's packets;
Output:
   N, the number of clusters inferred in by the algorithm;
   M = {C_1, ..., C_N}, the set of inferred clusters.
Begin   // the initial clusters' configuration
   N := m; M := φ;
   For i := 1 to N do
      C_i := {F_i};
      M := M ∪ {C_i};
   endfor;
   For i := 1 to m do
      Do
         modified := false;
         C := cluster that contains flow F_i;
         min := cost_function(C);
         For each Dst cluster, Dst ∈ M, Dst ≠ C do
            If cost_function(Dst ∪ {Fi}) < min then
               Move F_i from C to Dst;
               min := cost_function(Dst);
               If C is an empty cluster then
                  M := M − C; N := N − 1;
               endif;
```

```
            modified := true;
            break;
          endif;
        endfor;
      While modified = true;
   endfor;
end; // ClusterIdentification

Function cost_function(Cluster C):real is
   Initialize an empty vector T;
   For each flow F_i ∈ C do
      Append all values from vector T_i to vector T;
   endfor;
   Ascending sort vector T by its numerical values;
   Compute merged vector inter-packet arrival intervals Δ;
   Return standard deviation σ_C of Δ;
end; // cost_function
```

5 Algorithm Evaluation

The mechanism presented in the previous section has been tested on a simulated network having the topology shown in figure 3. The infrastructure of the simulated network is a real one, but the network's links operating at 100 Mbps had been shaped using HTB [10] in order to simulate lower link capacity. Each data source opens a TCP/IP connection to the destination and sends a continuous data flow over it. The data receiver records the incoming packets together with their arrival timestamps, for each incoming flow. The arrival times were recorder over a time interval of 160 milliseconds and are presented in figure 4.

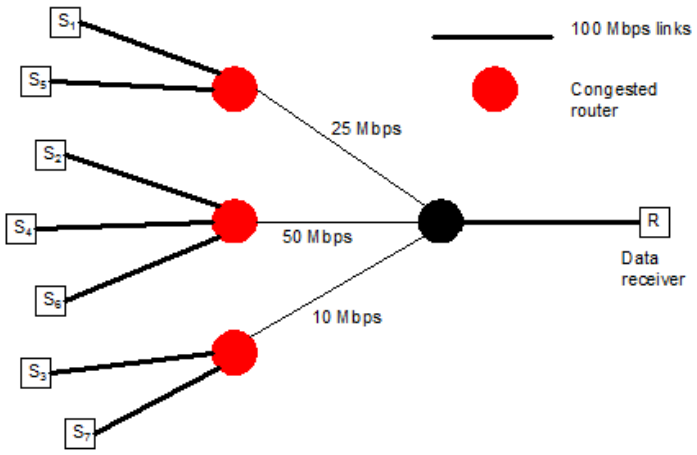


Figure 3: Simulated network

Flow	Arival times of flow's packets
F_1	7 31 55 63 85 93
F_2	4 12 22 24 38 43
F_3	10 44 56 106 117 148 159
F_4	19 26 40
F_5	15 22 39 46 70 77
F_6	7 9 15 30 33 36
F_7	21 32 68 81 94 128 138

Figure 4: Arrival times of flows' packets for the simulated network

Our method successfully detected three clusters of sources, correctly assigning the seven flows in three distinct macroflows, one macroflow for each congested link that the network contains. Figure 5 presents the identified macroflows, the flows that are part of each macroflow (cluster) and the equidistant packet arrival times for each macroflow.

Macroflow	Flows assigned to macroflow	Inter packet arrival times
C_1	F_1, F_5	7 8 7 9 8 7 9 8 7 7 8 8
C_2	F_2, F_4, F_6	4 3 2 3 3 3 4 3 2 2 4 3 3 2 2 3
C_3	F_3, F_7	10 11 11 12 12 12 13 13 12 11 11 10 10 11

Figure 5: Macroflows identified by our method for the simulated network

6 Conclusions and future work

In this paper we present a mechanism for better macroflow identification from the receiver point of view. This technique can be used inside an improved Congestion Manager to extend macroflows granularity.

The proposed mechanism successfully identified the correct macroflows when used on a simulated network. Future analysis must be performed in order to determine the performance of our method when applied to real bulk data collected from real traffic patterns. Methods for fast computation and implementation of the algortihm will be identified considering that the implementation is part of the very time-sensitive TCP/IP stack.

References

[1] S. Floyd, V. Jacobson, *Random Early Detection Gateways for Congestion Avoidance*, IEEE/ACM Transactions on Networking, 1(4), pp. 379-413, 1993.

[2] W. Stevens, *TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery*, IETF RFC 2001, January 1997.

[3] M. Allman, V. Paxson, W. Stevens, *TCP Congestion Control*, IETF RFC 2581, April 1999.

[4] H. Balakrishnan, S. Seshan, *The Congestion Manager*, IETF RFC 3124, June 2001.

[5] V. Paxson, *Measurements and Analysis of End-to-end Internet Dynamics*, Thesis Dissertation, 1997.

[6] A. Downey, *Using Pathchar to Estimate Link Characteristics*, Computer Communication Review, a publication of ACM SIGCOMM, volume 29, number 4, October 1999.

[7] D. V. Bufnea, A. Câmpan, A. S. Dărăbant, *Fine-Grained Macroflow Granularity in Congestion Control Management*, in Studia Universitatis, Vol. L(1), pp. 79-88, 2005.

[8] A. Câmpan, D. V. Bufnea, *Delimitation of Macroflows in Congestion Control Management Using Data Mining Techniques*, 4th ROEDUNET International Conference, Education/Training and Information/Communication Technologies - ROEDUNET '05, Romania, pp. 225-234, 2005.

[9] D. Katabi, I. Bazzi, X. Yang, *An Information Theoretic Approach for Shared Bottleneck Inference Based on End-to-end Measurements*, Laboratory for Computer Science, MIT, 2001.

[10] M. Devera, *Hierarchical Token Bucket Theory*, http://luxik.cdi.cz/~devik/qos/htb/manual/theory.htm, May 2002.

Darius Bufnea
"Babeş-Bolyai" University of Cluj-Napoca
Department of Computer Science
E-mail: bufny@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 206-211

Multimedia Visualisation for Breast Cancer

Yin Jie Chen, Simon Rajendran, Mark Tangney and Sabin Tabirca

Abstract: Multimedia represents a set of powerful tools for dissemination of data, in the context of teaching, public awareness education, and interpretation of research results. The Computing Resources for Research group at University College Cork is involved in the development of data visualisation systems for life science research. Building on the groups' skills in Multimedia, Information Technology and Learning and Cancer Research, we develop sophisticated multimedia content for researchers and practitioners in the life sciences. We have recently built a suite of animations on Breast Cancer aimed at patients, clinicians and researchers.
Keywords: Multimedia, Visualisation, Cancer, Breast Cancer

1 Introduction

Cancer is a class of diseases or disorders characterized by uncontrolled division of cells, and the ability of these cells to invade other tissues, either by direct growth into adjacent tissue through invasion or by implantation into distant sites by metastasis. Metastasis is defined as the stage in which cancer cells are transported through the bloodstream or lymphatic system. Cancer may affect people at all ages, but risk tends to increase with age. It is one of the leading causes of death in developed countries.[1]

Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple. The remainder of the breast is made up of fatty, connective, and lymphatic tissue.[2]

Breast cancer in women represents a significant health problem because of the numbers of individuals affected by this disease. Thirty percent of all cancers in women occur in the breast making it the most commonly diagnosed female cancer. The availability of easily interpretable information by way of multimedia, on early detection, treatment and prevention of breast cancer represents an extremely valuable resource to patients, researchers and clinicians.

Our application is a multimedia application presentsing general information for breast cancer in terms of biology, treatment and diagnosis. The application covers basic information for the public, in terms of basic breast cancer biology and education on early detection, and treatment, through to high level research concepts, molecular visualisation and data aimed at researchers and clinicians.

2 Multimedia and Visualisation

Multimedia

Multimedia is any combination of text, art, sound, animation, and video delivered to you by computer or other electronic or digitally manipulated means and is richly presented sensation [3]. Multimedia is media that uses multiple forms of information content and information processing (e.g. text, audio, graphics, animation, video, interaction) to inform or entertain the (user) audience [4]. The computer info can be represented through audio, graphics, image, video and animation in addition to traditional media (text and graphics).

Scientific Visualisation

Scientific and Information visualisation are branches of computer graphics and user interface design that are concerned with presenting data to users by means of images. Both fields seek ways to help users explore, make sense of, and communicate about data. They are active research areas, drawing on theory in information graphics, computer graphics, human-computer interaction and cognitive science.

Desktop programs capable of presenting interactive models of molecules and microbiological entities are becoming relatively common (Molecular graphics). The fields of Bioinformatics and Cheminformatics make heavy use of these visualization engines for interpreting lab data and for training purposes. The advent of such tools has resulted in the fields' biggest growth, growing contemporaneously with the internet.

Medical imaging is a huge application domain for scientific visualization with an emphasis on enhancing imaging results graphically, e.g. using pseudo-colouring or overlaying of plots. Real-time visualization can serve to simultaneously analyse results within or beside an analyzed (e.g. segmented) scan.

Multimedia in Scientific Visualisation

Multimedia technology is widely applied in scientific visualisation especially in medical area. Variable presentations are provided for scientific research. Scientific visualisation includes image processing, graphics, animation, 3D presentation and so on. Multimedia technology provides multiple approaches for visualization.

Microscopic images and illustrations are common presentations in the field of cancer visualization. Such imaging requires hi-tech, expensive equipment, and the outputs are device-specific, with interpretation requiring expertise.

New multimedia technology gives the cheap, easy to understand platform for cancer visualization especially for education. 2D animations can present complex biology processing from cell to whole body levels. . These animations can be embedded into computer applications, web pages and movies. 3D application can simulate object in different level with mathematic models.

3 Cancer Visualisations

In recent years there has been a great deal of interest concerning three catch-phrases: nonlinear dynamics, chaos, and complexity. This interest has led to a large number of popular-science articles relating to the visualization of cells, many of these feature very high end graphical simulations [5].

Technologies such as ultrasound are being used to evaluate solid tumours in 3D [6] an also 3D microscopy imaging to display 3D biological tissue architecture during carcinogenesis [7]. Models and Simulations are being widely used, for example Mathematical models have been used to describe ontogenetic growth [8]. Two-dimensional modelling has been used in the area of tumour growth and morphology [9], while simulations have been used in the area of benign avascular tumour growth [10].

Common breast cancer visualisations are static images and animations. The presentation formats are quite simple; some of contents are professional E.g. ultrasound images, charts and volume rendering 3D images that require expertise to interpret.

4 Problems

Cancer

Cancer is a group of diseases in which cells are aggressive (grow and divide without respect to normal limits), invasive (invade and destroy adjacent tissues), and often metastatic (spread to other locations in the body). These three malignant properties of cancers differentiate them from benign tumours, which are self-limited in their growth and do not invade or metastasize (although some benign tumour types are capable of becoming malignant).

Scope

Our visualization is focussed on breast cancer at body, tumour and cell levels. The application presents breast anatomy, breast cancer cell growth, tumour growth and spread (metastasis). The application details breast cancer examination and treatment also.

The main requirement is to present the location of tumour growth, cell alterations and spread in a realistic manner, and to educate on different treatment or examination approaches. To make these easy to understand and the application user interface easy to use, we need to create visual animations with audio and text, and a friendly simple user interface with interactions.

Usability

Usability is an attribute of the quality of a system.[11] Our application is for educational and research purpose. Therfore, both the contents and user interface must be easy to use and understand. The application has good flexibility for content extension also.

5 Solution

To handle multimedia content and user interface usability we selected Flash to implement application.

The Flash application is powerful in enabling interactive aspects in animations. The interactive aspect is very important to educational functions. These functions allow users to select content that they want to know, researchers could select content on micro level and others could select content on body or tumour level.

Flash content is easy to spread by Internet with web pages and its running requirement (web browser plug-in) is common. Graphics, audio, video and interactive components are easy to transfer by web in Flash and is therfore most suitable for educational purposes.

6 Application

Structure
Our application contains four main components:

- General Information
  This component contains information on breast anatomy and physiology. It highlights the main structures involved in lactation (milk-production) and cancer development. It also depicts the vascular (blood) and lymphatic drainage of the breast and axilla(Figure 2).

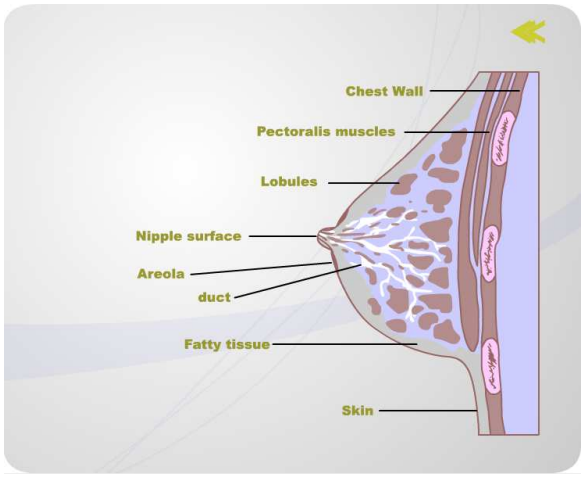


Figure 1: Screenshot of Breast Anatomy

- Cancer Biology
  This component features a detailed account on the evolution of breast cancer. It defines the different stages of cancer progression from in-situ, invasive and metastatic disease. The metastatic behaviour and cascade involving the vascular and lymphatic channels and distant spread are illustrated(Figure 3).

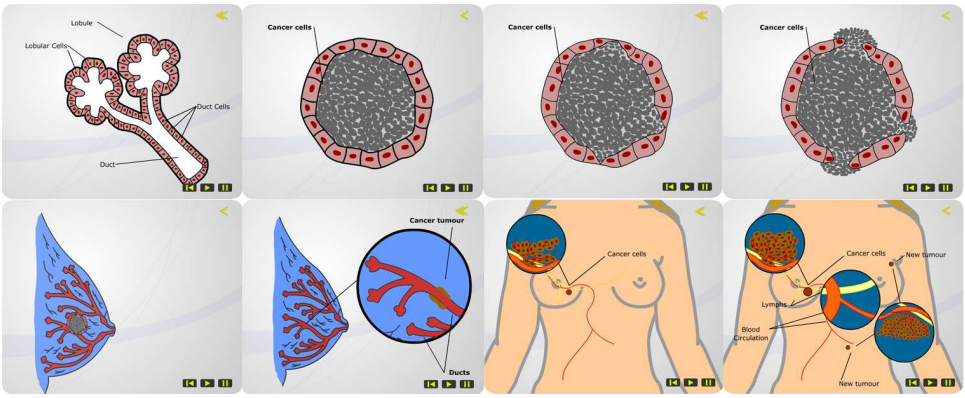


Figure 2: Screenshot of Cancer Biology

- Treatment
  This component highlights the current treatment strategies being used. It addresses the role of breast conservation surgery, sentinel node biopsy, chemotherapy, radiotherapy and hormonal therapy. It also features novel molecular and biological developments that may revolutionise breast cancer treatment in the future.

- Diagnosis
  This component outlines the symptoms and signs of breast cancer and the triple assessment used for breast screening. The role of current and new imaging modalities (mammography, ultrasound, magnetic resonance imaging (MRI)) and diagnostic procedures (fine needle aspiration (FNA), core biopsy) are explained. It also contains a patient video on performing a breast self-exam.

User Interface

The application provides a friendly and easy to use user interface(Figure 4). The contents are located into different category menus or submenus. The tree structure is easy to visit and transfer between them and easy to add new contents into. The user can control each animation or embedded movie by the use of some easy to use interactive tools. All controls can be done by mouse only.

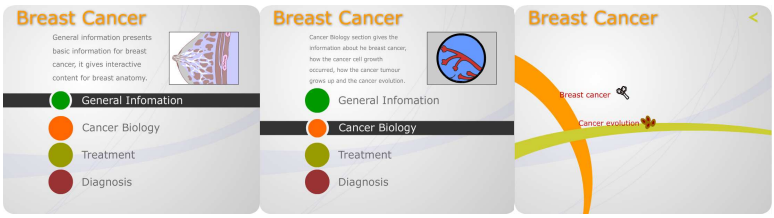


Figure 3: Application User Interface

Programming

We use ActionScript 3.0 to handle the programming in our application(Figure 5). ActionScript is the programming language for the Flash Player run-time environment. It enables interactivity, data handling, and much more in Flash content and applications. ActionScript 3.0 code can execute up to ten times faster than legacy ActionScript code[12].

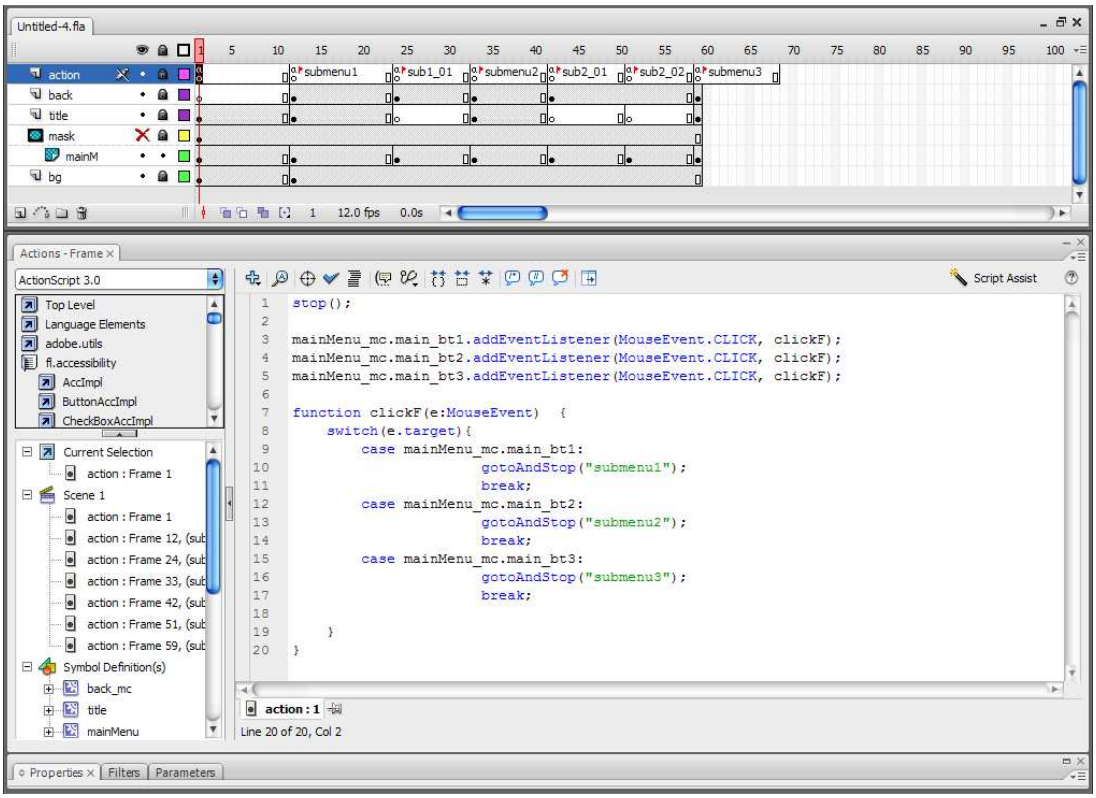


Figure 4: ActionScript in Application

The code shows how to handle the main menu buttons with event listener(Listing 1). "mainMenu.mainBt1" is

the name of button, the function "clickF()" handle the frame change when get a mouse click event. "submenu1" is the name of frame that contains sub menu.

```
mainMenu.mainBt1.addEventListener(MouseEvent.CLICK, clickF);
mainMenu.mainBt2.addEventListener(MouseEvent.CLICK, clickF);
mainMenu.mainBt3.addEventListener(MouseEvent.CLICK, clickF);
mainMenu.mainBt4.addEventListener(MouseEvent.CLICK, clickF);
function clickF(e:MouseEvent) {
 switch(e.target){
  case mainMenu.mainBt1:
            gotoAndStop("submenu1");
            break;
  case mainMenu.mainBt2:
            gotoAndStop("submenu2");
            break;
  case mainMenu.mainBt3:
            gotoAndStop("submenu3");
            break;
  case mainMenu.mainBt4:
            gotoAndStop("submenu4");
            break;
 }
```

Listing 1: Main menu scripting

This is the code for sub menus actions(Listing 2). "submenu2.sub2Bt1" and "submenu2.sub2Bt2" are options in a sub menu. Function "gotoAnimF2()" will change current frame when buttons get mouse click event.

```
submenu2.sub2Bt1.addEventListener(MouseEvent.CLICK, gotoAnimF2);
submenu2.sub2Bt2.addEventListener(MouseEvent.CLICK, gotoAnimF2);
function gotoAnimF2(e:MouseEvent) {
 switch(e.target){
  case submenu2_mc.sub2_bt1:
      gotoAndStop("sub2_01");
    break;
  case submenu2_mc.sub2_bt2:
      gotoAndStop("sub2_02");
    break;
 }
}
```

Listing 2: Sub menu scripting

7 Conclusion

Multimedia technology provides multiple approaches for scientific visualization especially for medical aspects. Our application involves these technologies to make a high usability system for professional and general use. This multimedia application can be an easy to use pattern for both education and research.

References

[1] Geoffrey M. Cooper, *The Cancer Book,* (ISBN:0867207701) Jones and Bartlett Publishers, 1993.

[2] American Cancer Society, *Breast Cancer Facts and Figures 2005-2006,* Atlanta: American Cancer Society, Inc, 2005.

[3] Tay Vaughan, *Multimedia: Making it Work, Seventh Edition* (ISBN:0072264519)McGraw-Hill Osborne Media, 2003.

[4] Wikipedia Multimedia, *http://en.wikipedia.org/wiki/Multimedia* Wikipedia

[5] M.H.F. Wilkinson, “Nonlinear Dynamics, Chaos-theory, and the "Sciences of Complexity" : Their Relevance to the Study of the Interaction between Host and Microflora”*Old Herborn University Seminar Monograph 10: New Antimicrobial Strategies*, pp. 110-130, 1997.

[6] M. Hunerbein and B.M. Ghadimi and S. Gretschel and P.M. Schlag, “Three-dimensional endoluminal ultrasound: a new method for the evaluation of gastrointestinal tumors”*Springer New York*, Vol. 24, Num. 5, pp. 445-448, 1999.

[7] arole J. Clem and Jean Paul Rigaut, “Computer simulation modelling and visualization of 3D architecture of biological tissues”*Springer Netherlands*, Vol. 43, Num. 4, pp. 425-442, 1995.

[8] G.B. West and J.H. Brown and B.J. Enquist, “A general model for ontogenetic growth”*Nature*, pp. 626, 2002.

[9] N. Adriana and Jose Mombach and M. Walter and F. deavila, "The interplay between cell adhesion and environment rigidity in the morphology of tumors"*Elsevier Science*, Vol. 322, pp. 546-554, 2003.

[10] E. L. Stott and N. F. Britton, "Stochastic Simulation of Benign Avascular Tumour Growth Using the Potts Model"*Elsevier Science*, Vol. 30, Num. 5, pp. 183-198, 1999.

[11] Donna Maurer, "What is usability?"*KM Column*, 2004.

[12] Flash CS3 Documentation, *http://www.adobe.com/support/documentation/en/flash/* Adobe

Yin Jie Chen, Simon Rajendran, Mark Tangney and Sabin Tabirca
University College Cork
Department of Computer Science, Cork Cancer Research Centre
Department of Computer Science, University College Cork, Ireland
E-mail: cyt1@cs.ucc.ie, simonrajendran@gmail.com, m.tangney@ucc.ie, tabirca@cs.ucc.ie

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 212-217

An Agent-Based Approach to Combinatorial Optimization

C. Chira, C.-M. Pintea, D. Dumitrescu

Abstract: Systems composed of several interacting autonomous agents are investigated for their potential to efficiently address complex real-world problems. Agents communicate by directly exchanging information and knowledge about the environment. Typical agent properties include autonomy, communication, learning, reactivity and mobility. It is proposed to further endow agents with stigmergic behaviour in order to cope with complex combinatorial problems. This means that agents are able to indirectly communicate by producing and being influenced by pheromone trails. Furthermore, each stigmergic agent has a certain level of sensitivity to the pheromone allowing various types of reactions to a changing environment. For better search diversification and intensification, agents can learn to modify their sensitivity level according to environment characteristics and previous experience. The resulting computational metaheuristic combines sensitive stigmergic behaviour with direct agent communication and learning to better addressing combinatorial optimization problems. The proposed model has been tested for solving various instances of NP-hard problems indicating the robustness and potential of the new metaheuristic.
Keywords: agent communication, stigmergy, sensitivity, multi-agent system, ant colony optimization

1 Introduction

Combinatorial optimization problems arise in many and diverse domains including network design, scheduling, mathematical programming, algebra, games, language technology, computational linguistics and program optimization. Metaheuristics are powerful strategies to efficiently find high-quality near optimal solutions within reasonable running time for problems of realistic size and complexity [1].

A metaheuristic combining stigmergic behavior and agent direct communication is proposed. The proposed model involves several two-way interacting agents. On one hand, agents are endowed with a stigmergic behavior [2, 3] similar to that of *Ant Colony Systems* [4, 5]. This means that each agent is able to produce pheromone trails that can influence future decisions of other agents. On the other hand, agents can communicate by directly exchanging messages using an *Agent Communication Language* - a behavior similar to that of multi-agent systems [6, 7, 8]. The information directly obtained from other agents is very important in the search process and can become critical in a dynamic environment (where the latest changes in the environment can be transmitted to other agents).

Furthermore, stigmergic agents are characterized by a certain level of sensitivity to the pheromone trail allowing various types of reactions to a changing environment [1].

Learning plays a crucial role in the functionality of the system. Agents can learn the environment characteristics and dynamically change the sensitivity level allowing a good balance between search exploration and exploitation.

The proposed model is tested for solving various instances of $\mathcal{NP}$-hard problems and numerical results indicate the potential of the proposed system.

The structure of the current paper follows the natural development process of the proposed model: description of stigmergic agents, introduction of agent sensitivity to pheromone trails and establishment of a learning mechanism for agents (presentation of the proposed computational model). Numerical experiments and results conclude the paper.

2 Stigmergic Agent Systems

Agents of the proposed model are able to communicate both directly and in a stigmergic manner using pheromone trails produced by agents. Communication in multi-agent systems [7] is necessary because agents need to exchange information or to request the performance of a task as they only have a partial view over their environment. Considering the complexity of the information resources exchanged, agents should communicate through an *Agent Communication Language (ACL)*.

Agents of the proposed model are able to exchange different types of messages in order to share knowledge and support direct interoperation. The content of the messages exchanged refers to environment characteristics and

partial solutions obtained. The information about dynamic changes in the environment is of significant importance in the search process.

Furthermore, the introduced model inherits agent properties such as autonomy, reactivity, learning, mobility and pro-activeness used in multi-agent systems [6, 7, 8]. The agents that form the system have the ability to operate without human intervention, can cooperate to exchange information and can learn while acting and reacting in their environment.

Agents are endowed with the ability to produce pheromone trails that can influence future decisions of other agents within the system. The idea of stigmergic agents was introduced in [1] where a system composed of stigmergic agents is outlined and illustrated by an example. Biology studies emphasize the remarkable solutions that many species managed to develop after millions of years of evolution. Self-organization [2] and indirect interactions between individuals make possible the identification of intelligent solutions to complex problems. These indirect interactions occur when one individual modifies the environment and other individuals respond to the change at a later time. This process refers to the idea of stigmergy [3].

The stigmergic behavior of agents is similar to that of the ants in the bio-inspired *Ant Colony Optimization (ACO)* metaheuristic [4, 5]. *ACO* simulates real ant behavior to find the minimum length path - associated to a problem solution - between the ant nest and the food source. Each ant deposits a substance called pheromone on the followed path. The decisions of the ants regarding the path to follow when arriving at an intersection are influenced by the amount of pheromone on the path. Stronger pheromone trails are preferred and the most promising paths receive a greater pheromone trail after some time.

3 Sensitive Stigmergic Agents

A robust and flexible system can be obtained by considering that not all agents react in the same way to pheromone trails. Within the proposed model each agent is characterized by a pheromone sensitivity level denoted by *PSL* which is expressed by a real number in the unit interval $[0,1]$. Extreme situations are:

- If $PSL = 0$ the agent completely ignores stigmergic information (the agent is 'pheromone blind');
- If $PSL = 1$ the agent has maximum pheromone sensitivity.

Small *PSL* values indicate that the agent will normally choose very high pheromone levels moves (as the agent has reduced pheromone sensitivity). These agents are more independent and can be considered environment explorers. They have the potential to autonomously discover new promising regions of the solution space. Therefore, search diversification can be sustained.

Agents with high *PSL* values will choose any pheromone marked move. Agents of this category are able to intensively exploit the promising search regions already identified. In this case the agent's behavior emphasizes search intensification.

4 Learning Sensitive Stigmergic Agents

Agents of the proposed model can learn to adapt their *PSL* according to the environment characteristics (and based on previous experience) facilitating an efficient and balanced exploration and exploitation of the solution space.

The initial *PSL* values are randomly generated. During their lifetime agents may improve their performance by learning. This process translates to modifications of the pheromone sensitivity. The *PSL* value can increase or decrease according to the search space topology encoded in the agent's experience. Low sensitivity of agents to pheromone trails encourages a good initial exploration of the search space. High *PSL* values emphasize the exploitation of previous search results. Several learning mechanisms can be engaged at individual or global level. A simple reinforcing learning mechanism is proposed in the current model. According to the quality of the detected solution, the *PSL* value is updated for each agent.

Agents with high *PSL* value (above a specified threshold τ) are environment exploiters and they will be encouraged to further exploit the search region by increasing their *PSL* value each time a good solution is determined. Agents with small

PSL value are good explorers of the environment and good solutions will be rewarded by decreasing agent *PSL* value (emphasizing space exploration). Let $PSL(A,t)$ denote the *PSL* value of the agent A at iteration t and $S(A,t)$

the solution detected. The best solution determined by the system agents (until iteration t) is denoted by $Best(t)$. The proposed learning mechanism works as follows:

Case 1: $PSL(A,t) > \tau$

- If $S(A,t)$ is better than $Best(t)$ then A is rewarded by increasing its PSL value according to the following learning rule:

$$PSL(A,t+1) = min(1, PSL(A,t) + exp(-PSL(t))/(t+1)2). \quad (1)$$

- If $S(A,t)$ is worse than $Best(t)$ then A is 'punished' by decreasing its PSL value according to the following learning rule:

$$PSL(A,t+1) = max(0, PSL(A,t) - exp(-PSL(t))/(t+1)2). \quad (2)$$

Case 2: $PSL(A,t) \leq \tau$

- If $S(A,t)$ is better than $Best(t)$ then A is rewarded by decreasing its PSL value according to the following learning rule:

$$PSL(A,t+1) = max(0, PSL(A,t) - exp(-PSL(t))/(t+1)2). \quad (3)$$

- If $S(A,t)$ is worse than $Best(t)$ then A is 'punished' by increasing its PSL value according to the following learning rule:

$$PSL(A,t+1) = min(1, PSL(A,t) + exp(-PSL(t))/(t+1)2).(4) \quad (4)$$

Agents learn the characteristics of the search space via a dynamic change in the PSL values. Good explorers of the solution space will be encouraged to more aggressively further explore the environment. Promising solutions already identified will be further exploited by rewarding the corresponding agent.

5 Learning Sensitive Stigmergic Agent System Model

The proposed model is initialized with a population of agents that have no knowledge of the environment characteristics. Each path followed by an agent is associated with a possible solution for a given problem. Each agent deposits pheromone on the followed path and is able to communicate to the other agents in the system the knowledge it has about the environment after a full path is created or an intermediary solution is built.

The infrastructure evolves as the current agent that has to determine the shortest path is able to make decisions about which route to take at each point in a sensitive stigmergic manner and based on learn information. Agents with small PSL values will normally choose only paths with very high pheromone intensity or alternatively use the knowledge base of the system to make a decision. These agents can easily take into account ACL messages received from other agents. The information contained in the ACL message refers to environment characteristics and is specific to the problem that is being solved. On the other hand, agents with high PSL values are more sensitive to pheromone trails and easily influenced by stronger pheromone trails. However, this does not exclude the possibility of additionally using the information about the environment received from other agents.

The algorithm of the proposed model is sketched below.

```
Algorithm Learning Sensitive Stigmergic Agent System
Begin
Set parameters
Initialize pheromone trails
Initialize knowledge base
While stop condition is false
Begin
 Activate a set of agents
 Place each agent in search space
 Do - For each agent
  Apply a state transition rule to incrementally build a solution.
  Determine next move (stigmergic strategy / direct communication)
  Apply a local pheromone update rule.
  Propagate learned knowledge.
 Until all agents have built a complete solution
 Update PSL value for each agent using proposed learning mechanism.
 Apply a global pheromone update rule
 Update knowledge base (using learned knowledge).
End While
End.
```

After a set of agents determines a set of problem solutions, the proposed model allows the activation of another set of agents with the same objective but having some knowledge about the environment. The initial knowledge base of each agent refers to the information about the path previously discovered by each agent.

6 Numerical Experiments

The proposed model has been tested for solving various instances of the well known $\mathcal{NP}$-hard Traveling Salesman Problem. This section presents the numerical results obtained for the *Generalized Traveling Salesman Problem (GTSP)* and *Asymmetric Traveling Salesman Problem (ATSP)*.

The proposed computational model allows agents to deposit pheromone on the followed path. Unit evaporation takes place each cycle. This prevents unbounded intensity trail increasing. The system is implemented using sensitive stigmergic agents with low initial *PSL* values.

The performance of the proposed model in solving *GTSP* is compared to the results of classical *Ant Colony System (ACS)* technique [9], the *Nearest Neighbor (NN)* algorithm, the GI^3 composite heuristic [10] and *Random Key Genetic Algorithm (rkGA)* [11]. The algorithm *Ant Colony System* for *GTSP* [9] is based on the *ACS* [4, 5] idea of simulating the behavior of a set of agents that cooperate to solve a problem by means of simple communications. In *Nearest Neighbor* algorithm the rule is always to go next to the nearest as-yet-unvisited location. The corresponding tour traverses the nodes in the constructed order. The composite heuristic GI^3 is composed of three phases: the construction of an initial partial solution, the insertion of a node from each non-visited node-subset, and a solution improvement phase [10]. The *Random Key Genetic Algorithm* combines a genetic algorithm with a local tour improvement heuristic. Solutions are encoded using random keys, which circumvent the feasibility problems encountered when using traditional *GA* encodings [11].

The data set of Padberg-Rinaldi city problems (*TSP* library [13]) is considered for numerical experiments. *TSPLIB* provides the optimal objective values (representing the length of the tour) for each problem. Comparative results obtained are presented in Table 1.

The proposed model gives the optimal solution for 7 out of the 10 problems engaged in the numerical experiments. For two other problems, the solutions reported by the proposed model are very close to the optimal value and better than those supplied by the other methods considered.

The proposed model for solving *ATSP* is implemented using sensitive stigmergic agents with initial randomly generated *PSL* values. Sensitive-explorer agents autonomously discover new promising regions of the solution space to sustain search diversification. Each generation the *PSL* values are updated according to the reinforcing learning mechanism described in Section 4. The learning rule used ensures a meaningful balance between search exploration and exploitation in the problem solving process.

The performance of the proposed model in solving *ATSP* is compared to the results of classical *ACS* technique and the *Max-Min Ant System (MMAS)* [12]. Several problem instances from *TSP* library [13] are considered for numerical experiments. Comparative results obtained are presented in Table 2. The parameters of the algorithm

Problem	Opt.val.	NN	GI^3	ACS	rkGA	Proposed Model
16PR76	64925	76554	**64925**	**64925**	**64925**	**64925**
22PR107	27898	28017	**27898**	27904.4	**27898**	**27898**
22PR124	36605	38432	36762	36635.4	**36605**	**36605**
28PR136	42570	47216	43117	42593.4	**42570**	**42570**
29PR144	45886	46746	**45886**	46033	**45886**	**45886**
31PR152	51576	53369	51820	51683.2	**51576**	**51576**
46PR226	64007	68045	**64007**	64289.4	**64007**	**64007**
53PR264	29549	33552	29655	29825	**29549**	29549.2
60PR299	22615	27229	23119	23039.6	22631	**22628.4**
88PR439	60099	67428	62215	64017.6	60258	**60188.4**

Table 1: Experimental results for solving Padberg-Rinaldi *GTSP* data set [13].

Problem	Opt.val.	ACS	MMAS	Proposed Model
Ry48p	14422	**14422**	**14422**	**14422**
Ft70	38673	38781	38690	**38682**
Kro124	36230	36241	36416	**36238**
Ftv170	2755	2774	2787	**2755**

Table 2: Experimental results for solving *ATSP*.

are similar to those of ACS: ten ants are used and the average of the best solutions is calculated for ten successively runs.

The proposed model detects a near-optimal or optimal solution for all the ATSP problems engaged in the numerical experiments. For one of the problem instances, all three methods compared find the optimal solution. For the other instances, solutions detected by the proposed model are very close to the optimal value and better than those supplied by the other methods considered.

The numerical experiments and comparisons emphasize the potential of the proposed approach to address complex problems and facilitate further connections between multi-agent systems and nature inspired computing.

7 Conclusions

Solving large complex problems - particularly those with a dynamic character - represents a challenging task. This paper takes an agent-based approach to solve combinatorial optimization problems. The components of a multi-agent system are endowed with a supplementary capacity - the ability of communication by environmental changes. Agents adopt a stigmergic behavior (being able to produce pheromone trails) to identify problem solutions and use direct communication to share knowledge about the environment. Agents can adapt various sensitivity levels to pheromone trails via learning. This approach results in a new metaheuristic able to address problems that involve very complex search spaces for which solutions are incrementally built by agents. Numerical experiments indicate the effectiveness and the potential of the proposed technique.

The primary elements of the proposed model include stigmergy, sensitivity and learning. Intelligent problem solutions can naturally emerge due to agent communication, autonomy and different levels of sensitivity to pheromone trails.

References

[1] C. Chira, C-M. Pintea, D. Dumitrescu, Sensitive Stigmergic Agent Systems, *Adaptive and Learning Agents and Multi-Agent Systems (ALAMAS)*, Maastricht, The Netherlands, 2 & 3 April 2007, MICC Technical Report Series 07-04 (Karl Tuyls, Steven de Jong, Marc Ponsen, Katja Verbeeck Eds.), pp. 51-57, 2007.

[2] S. Camazine, J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, E. Bonabeau, *Self organization in biological systems*, Princeton Univ. Press, 2001.

[3] P.-P. Grasse, La Reconstruction du Nid et Les Coordinations Interindividuelles Chez Bellicositermes Natalensis et Cubitermes sp. La Thorie de la Stigmergie: Essai d'interpretation du Comportement des Termites Constructeurs, *Insect Soc.*, 6, pp. 41-80, 1959.

[4] M. Dorigo, G. Di Caro, L. M. Gambardella, Ant algorithms for discrete optimization, *Artificial Life*, 5, pp. 137-172, 1999.

[5] M. Dorigo, C. Blum, Ant Colony Optimization Theory: A Survey, *Theoretical Computer Science*, 344, 2-3, pp. 243-278, 2005.

[6] H. S. Nwana, Software Agents: An Overview, *Knowledge Engineering Review*, 11, pp. 1-40, 1996.

[7] N. R. Jennings, An agent-based approach for building complex software systems, *Comms. of the ACM*, 44, 4, pp. 35-41, 2001.

[8] M. Wooldridge, P. E. Dunne, The Complexity of Agent Design Problems: Determinism and History Dependence, *Annals of Mathematics and Artificial Intelligence*, 45, 3-4, pp. 343-371, 2005.

[9] C-M. Pintea, C. P. Pop, C. Chira, The Generalized Traveling Salesman Problem solved with Ant Algorithms, *Journal of Universal Computer Science*, Graz, Springer-Verlag, in press.

[10] J. Renaud, F.F. Boctor, An efficient composite heuristic for the Symmetric Generalized Traveling Salesman Problem, *European Journal of Operational Research*, 108, pp. 571-584, 1998.

[11] L. V. Snyder, M. S. Daskin, A Random-Key Genetic Algorithm for the Generalized Traveling Salesman Problem, *European Journal of Operational Research*, pp. 38-53, 2006.

[12] T. Stutzle, H. H. Hoos, The Max-Min Ant System and Local Search for the Travelling Salesman Problem, *IEEE International Conference on Evolutionary Computation*, Piscataway (T. Back, Z. Michalewicz and X. Yao Eds.), IEEE Press, pp. 309-314, 1997.

[13] http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/

C. Chira, C.-M. Pintea and D. Dumitrescu
Babes-Bolyai University
Department of Computer Science
M. Kogalniceanu 1, 400084 Cluj-Napoca, Romania
E-mail: {cchira, cmpintea, ddumitr}@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 218-222

A Comparative Analysis on a Class of Numerical Methods for Estimating the ICA Model

Doru Constantin, Luminiţa State

Abstract: The reported work focuses on developing numerical methods for estimation the ICA model. Approximative schemes of the standard algorithm for estimating the independent components are developed using the secant method, combined method and successive approximations method. A comparative analysis on the efficiency of the above mentioned algorithms are reported in final section of the paper. The tests were performed for signals source separation purposes.
Keywords: Independent Component Analysis, Blind Source Separation, Numerical Methods

1 Introduction

An important problem arising in signal processing, mathematical statistical and neural networks is represented by the need of getting adequate representations of multidimensional data. The problem can be stated in terms of finding a function f such that the n dimensional transform defined by $s = f(x)$ possesses some desired properties, where x is a m dimensional random vector. Being given its computational simplicity, frequently the linear approach is attractive, that is the transform is

$$s = Wx \tag{1}$$

where W is a matrix to be optimally determined from the point of view of a pre-established criterion.

There are a long series of methods and principles already proposed in the literature for solving the problem of fitting an adequate linear transform for multidimensional data [1,4], as for instance, Principal Component Analysis (PCA), factor analysis, projection methods and Independent Component Analysis (ICA).

The aim of Independent Component Analysis is to determine a transform such that the components $s_i, i = \overline{1..n}$ becomes statistically independent, or at least almost statistically independent. In order to find a suitable linear transform to assure that (1) $s_i, i = \overline{1..n}$ become 'nearly' statistically independent several methods have been developed so far. Some of them, as for instance Principal Component Analysis and factor analysis are second order approaches, that is they use exclusively the information contained by the covariance matrix of the random vector x, some of them, as for instance the projection methods and blind deconvolution are higher order methods that use an additional information to reduce the redundancies in the data. Independent Component Analysis has became one of the most promising approaches in this respect and, consists in the estimation of the generative model of the form $x = As$, where the $s = (s_1 s_2, \ldots s_n)^T$ are supposed to be independent, and A is the mixing matrix $m \times n-$ dimensional of the model. The data model estimation in the framework of independent component analysis is stated in terms of a variational problem formulated on a given objective function.

One of the aims of the present reported research is to adapt the FastICA algorithm for developing numerical methods as well as to analyze the performances of the resulted algorithms from the point of view of signal recognition tasks.

2 The Versions of the FastICA Algorithm based on Numerical Methods

2.1 The standard FastICA Algorithm

In this section the ICA model and the standard FastICA algorithm are briefly exposed. The ICA model is state as $x = As$, where x is the observations vector and A is the mixing matrix of the original sources, s. The aim is to determine the sources, on the basis of x. One of the basic working assumption in estimation the ICA model is that the sources s are statistically independent and they have nongaussian distributions. This way the problem becomes to find the weighting matrix W (the demixing matrix), such that the transform $y = Wx$ gives suitable approximations of the independent sources.

In the following, the numerical estimation of the independent components is going to be obtained using the secant method, the combined method and the successive approximation approaches, the variational problem being imposed on the negentropy taken as criterion function.

The negentropy is defined by:

$$I(y) = H(y_{gauss}) - H(y) \tag{2}$$

where $H(y) = -\int p_y(\eta)\log p_y(\eta)d\eta$ is the differential entropy of the random vector y.

Being given that the Gaussian repartition is of largest differential entropy in the class of the repartitions having the same covariance matrix, the maximization of the negentropy (2) gives the best estimation of the ICA model. Although this approach is well founded from information point of view the direct use of the expression (2) is not computationally tractable, and some approximations are needed instead. We use the approximation introduced in (Hyvarinen, 98):

$$I(y) = [E\{G(y)\} - E\{G(v)\}]^2 \tag{3}$$

where G is an nonquadratic function, v and y are Gaussian variables of zero mean and unit variance. Some of the most frequently used expressions of G are,

$$G_1(y) = \frac{1}{a_1}\log cosh(a_1 y);\ 1 \le a_1 \le 2,\ G_2(y) = -\exp(-\frac{y^2}{2});\ G_3(y) = \frac{y^4}{4}$$

Note that the expressions of their first order derivatives are given by: $g_1(y) = \tanh(a_1 y);\ g_2(y) = y\exp(-\frac{y^2}{2});$ $g_3(y) = y^3$, respectively.

The variational problem can be formulated as a constraint optimization problem as follows,

$$\max F(w),\ \|w\|^2 = 1 \tag{4}$$

that is the objective function $F(w)$ has to be maximized on the unit sphere. In case the negentropy is taken as the objective function, we get,

$$F(w) = [E\{G(y)\} - E\{G(v)\}]^2 \tag{5}$$

where $y = w^T z$.

To solve the optimization problem the (5) the Lagrange multiplicators method can be used yielding to

$$F^*(w) = E\left\{zg(w^T z)\right\} - \beta w = 0 \tag{6}$$

where β is a real constant, $\beta = E\left\{w_0^T z g(w_0^T z)\right\}$ and w_0 is the optimum value of w.

The Newton method applied to (11) gives

$$w \leftarrow E\{zg\left(w^T z\right)\} - E\{g^{'}\left(w^T z\right)\}w \tag{7}$$

The weighting vectors being normalized, we arrive at the following approximation scheme,

1. $w \leftarrow E\{zg\left(w^T z\right)\} - E\{g^{'}\left(w^T z\right)\}w$
2. $w \leftarrow w/\|w\|$.

2.2 Adapted Versions of the FastICA Algorithm

The estimation of the ICA model can be obtained using numerical methods. In this sections, we proposed adapted versions of the FastICA algorithm based on the secant method, combined method and successive approximations method. Experimentally supported conclusions concerning the performance of these methods are reported in the next section.

Adapted FastICA Algorithm based on the Secant Method

A secant method [3] in solving the equation (11) yields to the following iterative scheme:

1. select the initial approximation w_0 and a randomly generated value a.
2. apply the updating rule: $\Delta w = -\frac{F^*(w)}{F^*(w)-F(a)}(w-a)$.
3. if the convergence criterion does not hold then go to the step 2, else take the last value of the w^k as the approximative solution of (11).

The convergence criterion is $\left\|w^{k+1} - w^k\right\| < \varepsilon$, where $\varepsilon = 10^{-N}$ is a small real constant, $N \in N^*$ given.

Adapted FastICA Algorithm based on the Combined Method

The method is a combination of the Newton method and the secant method and is represented by the following iterative scheme:

1. choosing the initial approximation w_0 and $\overline{w}_0$.
2. apply the Newton method: $\overline{w}_{n+1} = \overline{w}_n - \frac{F^*(\overline{w}_n)}{F^{*'}(\overline{w}_n)}$.
3. apply the secant method: $w_{n+1} = w_n - \frac{F^*(w_n)}{F^*(\overline{w}_n) - F^*(w_n)} (\overline{w}_n - w_n)$.
4. if the convergence criteria are not satisfied go to the step 2, else stop.

The initial approximation of the solution for the secant method is denoted by w_0 and for the Newton method is $\overline{w}_0$.

The convergence criterion is $\|w_n - \overline{w}_n\| < \varepsilon$, where $\varepsilon = 10^{-N}$ is a small given real constant, $N \in N^*$.

Adapted FastICA Algorithm based on the Successive Approximations Method

Using the successive approximation method, the approximations sequence can be written by $w = \varphi(w)$ where $\varphi(w) = w - {}^1\!/_M * F^*(w)$ and M is the maximum value of $F^{*'}$.

The updating rule becomes: $w \leftarrow w - {}^1\!/_M * [E\{zg(w^T z)\} - \beta w]$, where g is previously defined.

The Detailed Version of the FastICA Algorithm based on Secant Method, Combined Method and Successive Approximations Method

Using the previously methods approximations scheme together we get,

Step 1 : Center the data to mean.

Step 2 : Apply the whitening transform to data.

Step 3 : Select the number of independent components n and set counter $r \leftarrow 1$.

Step 4 : Select the initial guess of unit norm for w_r.

Step 5 : Apply the updating rules:

1. Case of Secant Method:
   $w_r \leftarrow w_r - \frac{F^*(w_r)}{F^*(w_r) - F(a)}(w_r - a)$, where a is initial choosing.
2. Case of Combined Method:
   $w_r \leftarrow w_r - \frac{F^*(w_r)}{F^*(\overline{w}_r) - F^*(w_r)} (\overline{w}_r - w_r)$, where $\overline{w}_r \leftarrow \overline{w}_r - \frac{F^*(\overline{w}_r)}{F^{*'}(\overline{w}_r)}$.
3. Case of Successive Approximations Method:
   $w_r \leftarrow w_r - {}^1\!/_M * [E\{zg(w_r^T z)\} - \beta w_r]$, where g is anterior defined and M is the maximum value of the function $F^{*'}$.

Step 6 : Apply the orthogonalization transform: $w_r \leftarrow w_r - \sum_{j=1}^{r-1} (w_r^T w_j) w_j$

Step 7 : $w_r \leftarrow w_r / \|w_r\|$.

Step 8 : If w_r has not converged ($\left\| {w_r}^{k+1} - {w_r}^k \right\| > \varepsilon$, where ε is a small real constant), go back to step 5.

Step 9 : Set $r \leftarrow r+1$. If $r \leq n$ then go to step 4.

3 Experimentally derived conclusions on the performance of the algorithms in case of signal mixtures

In this section we present a series of conclusions concerning the performances of the previously presented algorithms in determining the independent components in signal recognition. The absolute mean sum error (AbsMSE) is taken to measure the matching degree between initial signals and their restored versions, where

$$AbsMSE = \sum_{i=1}^{N} |s_i - s\_estimated_i| / N \tag{8}$$

s_i and $s\_estimated_i$ are the i-th pixel values in the initial and restored signals respectively, and N is the total number of pixels. The tests used the negentropy as a criterion function.

As a general conclusion the method based on successive approximations proved better recognition performances of the original signals as compared to the FastICA based on the secant method [3] or the combined method. In the following we briefly present two of our tests.

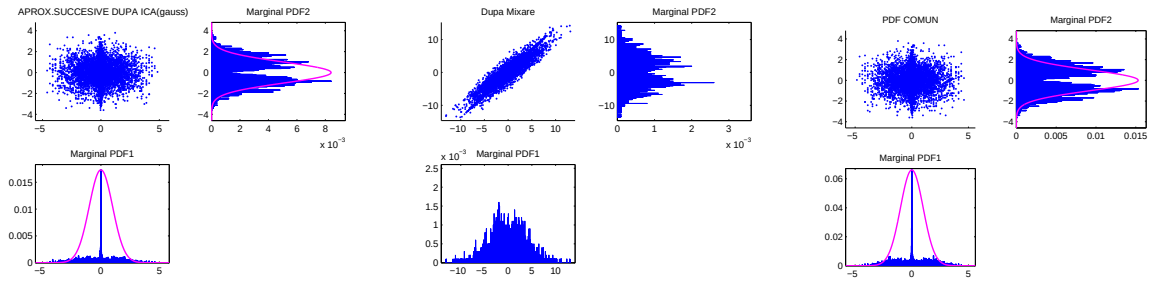


Figure 1: Source Signals Discovered by the Algorithm based on Successive Approximations (Left: 3 images), The Mixed Signals (Middle: 3 images) and Original Signals (Right: 3 images)

3.1 Test I

The observation data are given by two signals mixed and recorded on two independent components. The initial original sources were generated using the Matlab functions, and the results obtained by applying the algorithm based on successive approximations proved recognition performances comparable to the results obtained in case of standard FastICA method based on the Newton method and to the method FastICA method based on the secant method [3].

The source signals discovered by the algorithm based on successive approximations, the mixed signals and the source signals generated by Matlab subjected to the analysis procedure in independent components are depicted in figure 1, where the marginal densities corresponding to the two signals as well as the joint density which is common to the mixtures for the source signals discovered by the algorithm, for the mixed signals and for the source signals respectively, can be easily identified. The numerical results obtained by the application of the algorithm are summarized in table 1.

3.2 Test II

This test resembles the anterior test with the difference that it uses, as original signals, the uniform distribution signals. The figure 2 comprise the original source signals, the mixed signals and the source signals discovered by

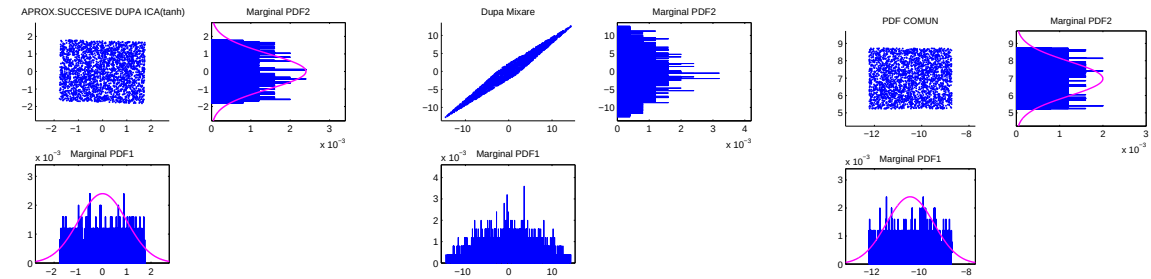


Figure 2: Source Signals (uniform) Discovered by the Algorithm based on Successive Approximations (Left: 3 images), The Mixed Signals (uniform) (Middle: 3 images) and Original Signals (uniform) (Right: 3 images)

the algorithm based on the successive approximations for the uniform signals case.

The results obtained after the comparative study regarding the methods used in the estimation of the ICA model, are similar to the ones from the first test conform with table 1.

Table 1: AbsMSE of versions of the FastICA Algorithm for experimental tests

FastICA	Test I	Test I	Test I	Test II	Test II	Test II
Basic Method	tanh (g_1) $\ldots * 10^{-2}$	exp (g_2) $\ldots * 10^{-2}$	kurt (g_3) $\ldots * 10^{-2}$	tanh (g_1) $\ldots * 10^{-2}$	exp (g_2) $\ldots * 10^{-2}$	kurt (g_3) $\ldots * 10^{-2}$
Newton	1.0692	1.0703	1.0688	2.4166	2.3166	2.4166
Secant	1.0726	1.0672	1.2785	2.3145	2.2754	2.3145
Combined	1.0746	1.0689	1.1268	2.3362	2.2938	2.3682
Successive Approx.	1.0666	1.0665	1.0668	2.2985	2.2981	2.2984

4 Summary and Conclusions

In this article we developed numerical methods for estimation of the ICA model. First, we have deducing the ICA algorithm based on tangent method using as objective function the negentropy, then we formulate the iterative scheme for the secant method, combined method and successive approximations method. In final we establish a recognition performance of the proposed algorithms in signal applications.

References

[1] Hyvarinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis*, John Wiley & Sons, (2001).

[2] Hyvarinen, A., Oja, E.: "Independent component analysis: algorithms and applications", *Elsevier Science*, Neural Networks **13** (2000) 411–433.

[3] Cho, Y.H., Hong, S.J., Park, S.M.: "Face Recognition by Using Combined Method of Centroid Shift and Fixed-Point ICA Based on Secant Method", *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*, IEEE Computer Society, (2006).

[4] Stone, J.V.: *Independent Component Analysis*, The MIT Press, Cambridge, Massachusetts London, England, (2004).

[5] Inki, M.: *Extensions of Independent Component Analysis for Natural Image Data*, Helsinki University of Technology, Espoo (2004).

[6] State, L., Cocianu, C., Panayiotis, V., Constantin, D.: "PRINCIPAL DIRECTIONS - BASED ALGORTHM FOR CLASSIFICATION TASKS", *Proceedings of the Ninth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2007) 26-29 September 2007 in Timisoara*, IEEE Computer Society.

[7] Hyvarinen, A.: "Survey on Independent Component Analysis", *Neural Computing Surveys*, Helsinki University of Technology, Finland, **2** (1999) 94-128.

[8] Roberts, S., Everson, R.: *Independent Component Analysis: Principles and Practice*, Cambridge University Press, (2001).

Doru Constantin, Luminiţa State
University of Piteşti
Department of Computer Science
Address Str. Tg. din Vale, Nr.1
E-mail: cdomanid@yahoo.com, radus@sunu.rnc.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 223-227

Simulation Model Linked to a Knowledge Based System for Evaluating Policies in the Operation of an Underground Mine

F.M. Cordova, L. Canete, L.E. Quezada, F. Yanine

Abstract: This paper presents a Simulation Model linked to a Knowledge Based System (SM-KBS) built to support the scheduling for an underground mine. Both systems are designed by means of hierarchical, colored and temporal Petri Nets. Simulation Model simulates the operation of the production, reduction and transport levels. Knowledge Based System is activated by events produced in daily operations and yields the results of registered events and the actions taken to solve the problem, generating operation rules. The proposed model, parametrical and scalable, allows different types of mine operations and scenarios with the objective of obtaining data for decision-making. It helps to evaluate different policies against specific scenarios for programming the activities in the mine seeking to enlarge the equipment productivity. In addition, the model allows the feasibility assessment of the Daily Master Plan based on the input data of the simulation model.
Keywords: Simulation Model, Cognition and Control, Knowledge Based System, Petri Nets, Underground mining.

1 Introduction

During daily operations of Codelco's underground mine Teniente 4 Sur (Chile) there are continuous events being generated which alter the normal work cycle of the mine, affecting its activities in their diverse levels or resources [1]. Observation and analysis of their behavior and the effects of these events thereafter in daily main operations is of the most importance. It is in this context, that it is necessary to simulate the behavior of an underground mine, whose continually functioning does not allow to be available for its study, thus providing in this way new alternatives to improve its productivity, its efficiency and efficacy.

Petri Nets allow modeling the production process of the mine because it is possible to built various independent modules and then to produce a given configuration as a combination of those modules [2]. Petri Networks are being used for modeling dynamic operation of discrete systems, mainly in manufacture [3], [4],[5]. They are also utilized like a very useful tool for modeling, to analyze, to simulate and to control production systems [5],[6],[7]. In this way, a number of basic networks representing different elements of the mine are built and then they are combined to represent the hole system. Hierarchical and temporal Petri Nets have allowed the construction of of a simulation model with a knowledge based system built in, which is parametric, scalable and adaptable to various configurations of underground mines. In Petri Nets, each element represents a characteristic feature that is present in any system to be modeled and operations are based on three basic elements: places, transitions and tokens.

2 Organization of the production system

In the "El Teniente" mine processing cycle, the rocks in higher levels are reduced in size. However, the transition from secondary rock to primary rock creates a problem of inefficiency causing a fall of productivity up to 10%. This situation forces the company to use a mechanized method of exploitation called "Block Caving" which is shown in Figure 1. The production system consists basically of three main levels. At the production level the rocks are taken 18 meters down the ground level to the ore-passes. Then, the material is taken out by load-haul-dump (LHD) vehicles, transported in streets and dumped in the pits. At the reduction level located 35 meters below the production level, the mineral falls from pits into rock-breaker chambers. Here a rock-breaker reduces its granulation to less than a cubic meter and then it follows to the transportation chimneys (Mailbox). At the transportation level, the mineral is loaded from the mailboxes to the train (Crossed).

3 Simulator and KBS integrated in a single model

A single model integrates a Simulation Model (SM) and a Knowledge Based System (KBS). The simulator, which is being fed by the Daily Master Plan, allows processing and simulation of the different scenarios that may

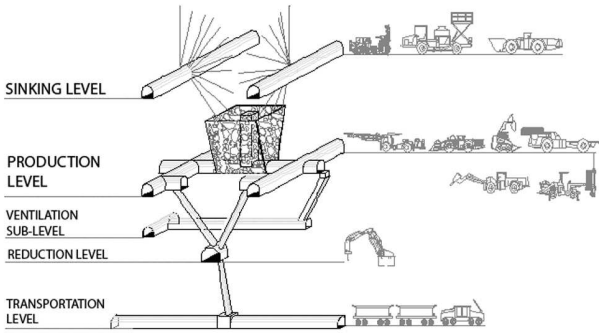


Figure 1: Block Caving Exploitation Method.

take place during daily mine operations, utilizing as a basis the historical data of the different shifts. KBS provides the knowledge acquired by experts on the operation of all resources that participate in the productive process, as well as the possible flaws of those resources. This allows to generate events, solutions and new operation rules for the system identifying critical failures of the system. The architecture of the Simulator linked to KBS is presented in Figure 2. The Daily Master Plan defines the operation for each available resource according to the production goals, also indicates the extraction points, tons. to be extracted, availability of pits and ore-passes. Simulated scenarios allows the simulation of a working shift in the mine under operational conditions without reprogramming. The system provides recommendations regarding actions to be taken, whenever some unexpected event happened. Reports will reflect a summary of the outcomes and falls which occurred, also recommendations delivered by the system and the final actions taken in the process. Reprogramming orders will be executed whenever a major outcome is registered and the needs to reprogram activities to achieve the daily goals are detected. Generation of new operational rules is activated when some outcome is registered revealing no final solution. Events or outcomes that perturb the established working program can force to modify it to achieve the production goals.

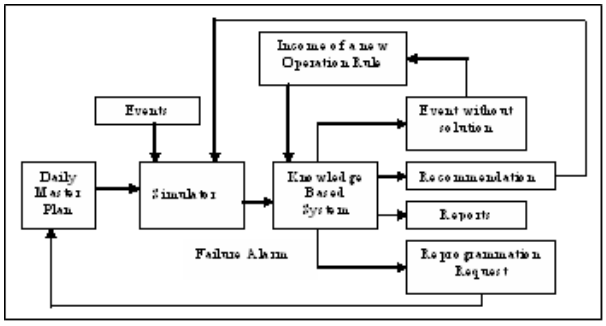


Figure 2: Smulator and KBS.

3.1 The Simulation Model(SM)

The proposed model describes the operation of load-haul-dump vehicles (LHDs) at the production level, also the operation of rock-breakers equipment at the reduction level from the beginning of a shift up to its end. LHDs vehicles moves through the tunnels inside the mine according to a given schedule, which will be affected by different events such as own breakdowns, rock-breakers breakdowns, etc. It is comprised of three sub-modules: production, reduction and transport. The production level module shown in Figure 3 initiates itself with the place denominated Generator LHD, from which tokens are activated which represent the LHD resources that contain the start up attributes of the simulation, yielded by the Daily Master Plan. Arriving to the street module, the token must identify the action that it will realize (entering or leaving street) once the token has continued on the indicated route, it must enter the first physical resource fixed in the mine and of the level production denominated Street (C1R to C15R, C1l to C15L).

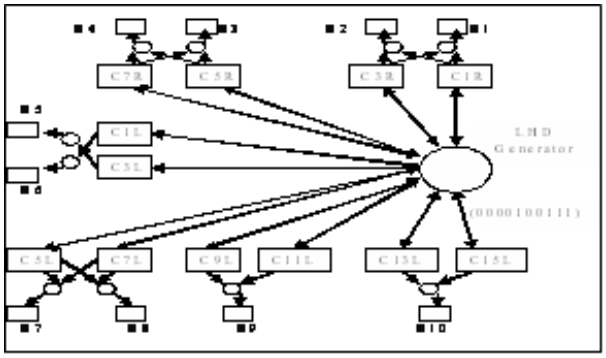


Figure 3: Production Level Module.

The Street Module has three modular structures: Ore-pass Module, Two Ore-passes Module and Two Ore-passes-Pit Module, which communicate with the rest by means of a Routing Tree. The resources at the Reduction Level are the Rock-breaker and the Rock-breaker operator. These resources are shared by a pair of streets and a design of the connection of production/reduction level is done. In the Transport Level the material is loaded and kept in Mailboxes, until it reaches its limit of storage capacity (280 tons.). In the case the resource Mailbox presents any failure, the Mailbox Operator must aid and repair it. In the model, each one of the LHD bucket represents a token therefore if a Mailbox concentrates 40 tokens, a new token will be generated in the Crossed. Figure 4 shows the connection between reduction and transport level including failure generation and solution given by KBS. The Routing Tree Module of the LHD has the function of directing the token of LHD through each one of the modules previously defined, assigning to each token that enters and leaves the master Daily Plan a unique vector of nine attributes (a b c d e nb ca l) which contain a unique combination representing a unique movement to be followed by the modules that must pass through.

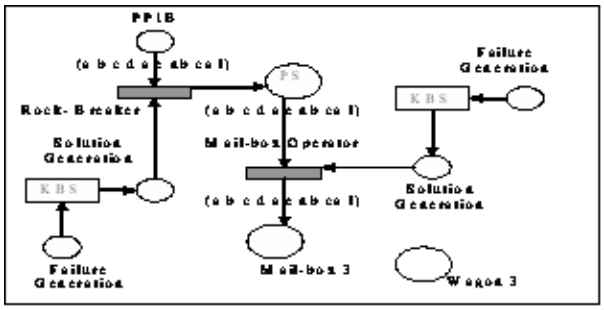


Figure 4: Reduction-Transport Module.

3.2 The Knowledge Based System (KBS)

KBS is the engine of knowledge which allows the simulation module to include solutions to the flaws that might have happened at each level. The KBS module integrates Production, Reduction, Transport and Human Resources (HH RR)levels allowing separation of failures that will happen as operations go along. KBS module is activated by a token associated to some particular input event. All token activated have some specific attribute "z" which describes the characteristics of the event. This token generation was defined as a unform probability which helps modeling the events. Once the event has been identified together with its solution, it will leave the KBS giving recommendations related to one of fifty six (56) failures that have been modeled. The events that happen at each level are associated to it frequency of occurrence per shift. According to this frequency a numerical interval is associated to each level, resource or failure. Each token generated in the initial transition of the model is characterized by numerical values. A first evaluation identifies the level affected by the input event at the production, reduction, transport and human resources level. A second evaluation identifies the resources affected by the events at each level. Six modules have been built at the production level (LHD, Street, Ore-pass, Pit, mini

LHD) as shown in Figure 5. The transitions have specific conditions that identify if the token fulfil or not the attributes defined by the condition, according to the corresponding histogram showed in Figure 5. The transitions include actions that allow the assignment of new features to the accepted tokens.

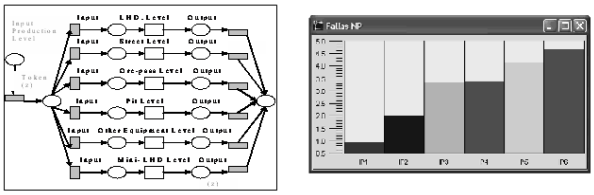


Figure 5: Production Module and Failure Histogram.

A third evaluation at the production level identifies the failure associated to specific resources. The LHD is a production level resource which is continuously exposed to faults. The potential events which have been defined for this resource shown in Figure 6 are: repair, major failure, minor flaw and oil supply. This design also includes a module which considers all possible resource failures at the reduction level and at the transport level which affect the system operation. The resources considered are: rock-breaker equipment, rock-breaker operator, mailbox, mailbox operator and crossed (trains).

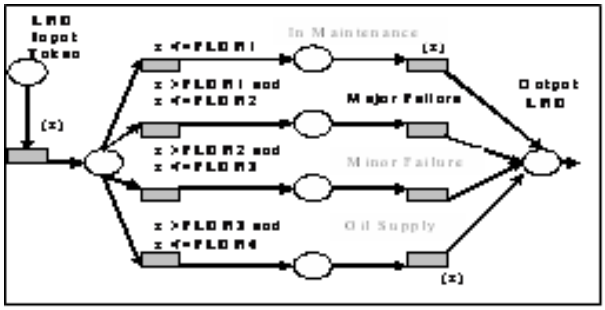


Figure 6: LHD Module.

4 Results

The SM-KBS was implemented using Pace Petri Nets Tool using different scenarios in a shift for evaluating the operation. Several iteration were done generating between 8 and 13 random failures which affect the resources of the system, altering the initial programmed production. In each case, responses and recommendations were generated, as well as real time operation graphs versus the number of flaws and real tonnage versus programmed tonnage (Figure 7). The system can identify the critical failures for the programmed operation, utilizing colored Petri Nets. These kind of failures implies a reprogramming of the operation. KBS calls for the flawed resource to reprogram its operation. The failures associated to the human resource are significants since its failure impacts all the other resources that require of its participation for the operation. The most frequent failures were related with the production level (LHD with minor failure) and in the transport level (Crossed with minor failure).

For the case of LHD with minor failure, the recommendation is: call the mechanics for them to do the repair on site. In the case of Crossed failure with a minor failure, the recommendation is: call a task team to carry out the necessary arrangements. When an undetermined failure is found, identifying the level, or levels affected, its frequency, the resources involved, and the estimated result of this,a new operation rule is added to the KBS.

5 Summary and Conclusions

The Simulator Linked to Knowledge Based System allows to study the real operation of an underground mine, including the generation of events affecting it. The design covers the three operational levels of the mine, including

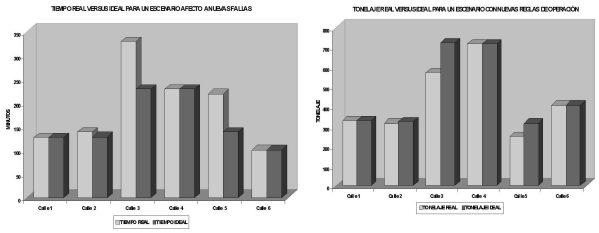


Figure 7: Real time versus the programmed time and real versus programmed tonnage.

human resources events. The use of hierarchical, colored and temporized Petri Nets in the design gave more flexibility to the model and allowed a hierarchy of the levels with their resources. In order to identify the failures or critical events that take place in the operation of the mine, the colored Petri Nets were used, tool that helped to give the aspect of critical that they have in the real operation. The programming considers 56 types of failures and it is possible to modify the probability of occurrence of one of them. Regarding the results obtained, the model was utilized to investigate the impact of the different types of failures. It was found that, in order of importance, the events affecting the production are the blocking of a secondary ore-pass, a minor breakdown of a rock-breaker and a major breakdown of a LHD vehicle. The effect of human resource failure is global and it affects directly he amount of production for the shift. When failures occurred, the KBS gave the correct solution that was programmed. The delay that was originated for each failure was scheduled according to the information given by experts in terms of time consumed in each case. The incorporation of critical failures in the modeling process is another innovative characteristic which represents in a realistic way the operation of the mine. For outcomes which do not generate a recommendation or solution defined in the KBS, a new outcome as a case is generated, adding new knowledge or a solution which generates new operational rules. These results are useful to the managers, because they know now in which resources to pay attention and improve the maintenance of those resources.

References

[1] Atero L.R., Cordova F.M., Quezada L.E., Olivares V., Sepulveda J., "Conceptual Model of Virtual Supervising Operation System VOSC," *Proc. of the 17 International Conference on Production Research, Virginia, USA*, 2003.

[2] Ghaeli M., Bahri A. B., Lee P. and Gu T., "Petri-net based formulation and algorithm for short-term scheduling of batch plants," *Computers and Chemical Engineering*, Vol. 29, Issue 2, pp. 99-102, 2005.

[3] Gu T., Bahri A.B., "A survey of Petri Nets applications in batch manufacturing," *Computers in Industry*, Vol. 47, Issue 1 pp. 99-102, 2002.

[4] Ounnar F., Ladet P., "Consideration of machine breakdown in the control of flexible production systems," *International Journal of Computer Integrated Manufacturing*, Vol. 17, Issue 1, pp. 62-82, 2004.

[5] Proth J.M., Xie X., *Petri Nets: A Tool for Design and Management of Manufacturing System,* John Wiley & Son Inc, 1997.

[6] Rosell J., "Assembly and Task Planning using Petri Nets: A Survey," *Proceedings of the Institution of Mechanical Engineers-Part B- Engineering Manufacturing*, Vol. 18, Issue 10, pp. 987-994, 2004.

[7] Zhang W.,Freiheilt Th., Yang H., "Dynamic Scheduling in Flexible Assembly System based on timed Petri Nets Model," *Robotics and Computer Integrated Manufacturing*, Vol. 21, Issue 6, pp. 550-558, 2005.

Felisa M. Cordova, Lucio Canete, Luis E. Quezada, Fernando Yanine
University of Santiago of Chile, USACH
Department of Industrial Engineering
Ecuador 3769, Santiago, Chile
E-mail: fcordova@usach.cl

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 228-231

Solving Fuzzy TSP with Ant Algorithms

Gloria Cerasela Crişan, Elena Nechita

Abstract: In this paper we describe the first Ant algorithm designated for fuzzy TSP (FTSP). Our work consists of two parts. The former part is dedicated to FTSP specification. The latter transforms the Ant System (chronologically the first Ant algorithm) in order to tackle the FTSP, and implement the new algorithm on a FTSP instance.
Keywords: Ant algorithm, Combinatorial Optimization, Fuzzy set.

1 Introduction

Inspired by the ants's behaviour, the Ant algorithms benefit from a progressive theoretical support and they are applied to many problems from engineering, economy, shipping, and communications.

When moving on the ground, real ants lay a substance called pheromone, tracing their walk. It has been observed that individuals which, afterwards, pass by the marked territory detect the trace and prefer the marked path instead of the unmarked zones.

The algorithms that solve problems using the ant colony model are called Ant algorithms. Marco Dorigo presented the Ant System (AS) in 1992 in his PhD thesis [1]; AS is the first Ant algorithm designated for Traveling Salesman Problem (TSP).

The real life problems deal with imperfectly specified knowledge. This means that some degree of imprecision, uncertainty or inconsistency is embedded in the problem specification. The well-founded theory of fuzzy sets is a special way to model the uncertainty : "More often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership. For example, the class of animals clearly includes dogs, horses, birds, etc. as its members, and clearly excludes such objects as rocks, fluids, plants, etc. However, such objects as starfish, bacteria, etc. have an ambiguous status with respect to the class of animals.". [6]

We selected the study of fuzzy TSP, as no approach has been made using Ant algorithms. The only fuzzy variant approached with Ant algorithms is fuzzy JSP [4].

In Section 2 we define the fuzzy TSP (FTSP), starting from TSP. In Section 3 we define the fuzzy AS (FAS), an Ant algorithm for FTSP. Section 4 describes the results of FAS on fuzzy variant of *lu980*, a TSP instance from [8].

2 Fuzzy TSP

TSP is a well-known combinatorial optimization problem, often representing the class of NP-hard problems. In the graph-theory formulation, it seeks in a complete weighted graph for a minimum weight Hamiltonian cycle.

In FTSP, the classic distances between nodes d_{ij} are substituted by fuzzy triangular numbers (m_{ij}, d_{ij}, p_{ij}) [3], defined as it follows.

Definition 1. *Given a complete weighted graph with n nodes and distance matrix* $(d_{ij}), 1 \leq i, j \leq n$, *and a parameter* $\lambda \in [0,1)$, the fuzzy distance matrix *is made up by triangular numbers* $((m_{ij}, d_{ij}, p_{ij})), 1 \leq i, j \leq n$, *where* m_{ij} *and* p_{ij} *are independent random positive values, smaller than* λd_{ij}.

Definition 1 introduces fuzzy distances, governed by the parameter λ. Basically, λ sets the maximum length of the interval that extends the numerical value d_{ij} from the deterministic, classic TSP. This parameter could be seen as a "fuzzyfication degree": if $\lambda = 0$, then we are in the classic TSP case; if λ is close to 1, then the uncertainty degree is very high, as the intervals $(d_{ij} - m_{ij}, d_{ij} + p_{ij})$ are very large.

Definition 2. *Given a complete weighted graph with n nodes and fuzzy distance matrix* $((m_{ij}, d_{ij}, p_{ij}))$, $1 \leq i, j \leq n$, the length *of a path is the sum of the fuzzy distances on path's edges.*

Definition 2 is consistent, as the set of triangular numbers is closed under addition. Unfortunately, not every two triangular numbers could be compared. And so, minimum of two such numbers does not always exist. Our choice is to implement a total, random operator described by the following definition.

Definition 3. *Given a complete weighted graph with n nodes and fuzzy distance matrix* $((m_{ij}, d_{ij}, p_{ij}))$, $1 \leq i, j \leq n$, *when the lengths of two paths could not be compared, one of them is randomly chosen as the* shorter *path.*

We are now able to define the fuzzy TSP.

Definition 4. *Given a complete weighted graph as in Definition 1*, the FTSP *seeks for a Hamiltonian cycle with the least weight, computed using Definitions 2 and 3.*

At the end of this section, there are some remarks to be made.

1. Definition 3 introduces a new level of randomness; coupled with the fuzziness from Definition 1, we are now in the presence of two types of uncertainty. This is a very interesting approach, but we have to mention that the random operator has an unpredictable short-time behaviour.
2. Of course, there are many other methods of fuzzyfication; all we had to do was to pick one. So, there are other ways to define the FTSP. The reasons why we use triangular numbers are: the set of triangular numbers is closed under addition and so under the averaging operator; the comparison is easy to make; the implementation as a data structure with three components is also easy to do; and (the most important) they suitably model the real situations when the edges' costs are uncertain due to some specific conditions (i.e. bad weather).
3. The result is a triangular number. We believe that its study (support, modal point, center of mass) could reveal interesting properties, that will be lost by standard defuzzyfication.

3 Fuzzy Ant System

Ant algorithms are stochastic and heuristic methods for solving optimization problems. Their nature leads to very-good, near-optimum solutions with affordable cost of resources (usually, computing time). Intense researches are dedicated both to theoretical study of their properties, and to empirically adapt them to specific problems or problem types [2].

The first Ant algorithm was Ant System (AS), designated to heuristically solve the TSP. The problem representation is the total weighted graph with n nodes, where a Hamiltonian cycle with minimum length is searched. In AS, some time-increment, starting from 0, is to be defined. Each artificial ant is an agent that chooses the next node with a probability reflecting the distance and the pheromone quantity on the edge. The already visited nodes are forbidden, using an agent's private list. After all agents finish their cycles, the pheromone values are updated. After that, the shortest current path is compared with the global-shortest path. If a shorter path is found at the current iteration, this path updates the global-shortest path. The next step erases the forbidden lists, the agents are re-located in their starting positions, and the algorithm is started again, until it reaches a pre-established value for the *NC* parameter (restarting counter). Other parameters are: α, β (balance the colony's history and the heuristic information), Q (the scale of the new pheromone values), ρ (the evaporation degree for the pheromone trace), and m (the constant number of artificial ants).

In order to tackle the FTSP, we have to slightly modify the AS:

1. Replace the distance matrix with a matrix of data structures, each composed by 3 scalar values.
2. Implement the addition described by Definition 2.
3. Implement the comparison described by Definition 3.

Definition 5. *The AS modified according to the above descriptions is called* fuzzy AS (FAS).

The FAS is an algorithm that works with fuzzy numbers, implemented as data structures. We consider that FAS is a natural extension of Ant algorithms for deterministic problems.

Given a fuzzy problem, there are two other methods to solve it:

1. Defuzzyfication (by replacing the fuzzy values with numbers, obtained by one of the classic defuzzyfication methods - the center of mass method is the most known), followed by the classic algorithm execution. This "reductionist" method is not suited for uncertainty research, as already mentioned at the end of Section 2.
2. Fuzzy algorithms [5] ("an ordered set of *fuzzy instructions* that upon execution yield an approximate solution to a given problem" [7]), an extremely new class of algorithms, as far as we know, not yet applied to optimization problems.

Our choice is an intermediary one, as it uses classic algorithms, modified in order to manipulate data structures, that model fuzzy numbers.

4 Implementation and Results

We have implemented the FAS algorithm on the FTSP instance based on the Euclidean instance *lu980* [8]. For the deterministic case, the optimum integer value is known: 11340. The application is written in Java and executed 5 times on 1.47 MHz processor with 256 MB RAM, under Windows OS. The results are averaged on integer values. The other parameter values are: $\alpha = 1, \beta = 5, Q = 100, \rho = 0.67, NC = 100, m = 50$.

In Table 1 are presented the results for $\lambda \in \{0.01, 0.05, 0.1, 0.2\}$.

λ	Triangular number
0.01	(71, 13934, 68)
0.05	(402, 14378, 384)
0.1	(917, 15720, 884)
0.2	(1472, 14587, 1392)

Table 1: Results for fuzzy variant of *lu980*

As expected, when λ increases, the results have supports with increasing length. This means that the fuzzyfication degree is directly reflected in the length of the results' intervals.

The modal points do not exhibit a pattern. An explanation could arise from the random comparison method from Definition 3. This unexpected behaviour shows that the fuzzyfication degree is not directly reflected in the result's modal points. The standard defuzzyfication using the modal points is inappropriate in this case.

Following from the FTSP definition, we do not know the optimum value of the fuzzy variant of *lu980*. Moreover, the FTSP instance has different values for the distances, at each FAS run; so at each run, there are different optimum values. This means that from the classical point of view, *FTSP does not have optimum value*.

5 Summary and Conclusions

This paper presents a fuzzy variant of TSP (FTSP), with triangular numbers as weights. The support of these triangular numbers is governed by a variable called the fuzzyfication degree (λ). The FTSP is heuristically solved using a modified Ant algorithm, called FAS. The results on a FTSP instance show that the fuzzyfication degree is directly reflected in the lengths of the intervals, but not on the modal points.

Future work will investigate other fuzzyfication methods, the influence of the randomness in the solutions' quality, other FTSP instances (Euclidean, non-Euclidean, asymmetric, etc.). The theoretical investigation of the concept of optimum solution for FTSP is extremely important, as it could start the study of theoretical properties as convergence or invariance.

References

[1] M. Dorigo, "Optimization, Learning and Natural Algorithms", PhD Thesis, Department of Electronics, Politecnico di Milano, Italy, 1992.

[2] M. Dorigo, T. Stutzle, *Ant Colony Optimization*, MIT Press, Cambridge, 2004.

[3] G. Klir, T. Folger, *Fuzzy sets, uncertainty and information*, Prentice Hall, New Jersey, 1988.

[4] J. Montgomery, C. Fayad, S. Petrovic, "Solution Representation for Job Shop Scheduling Problems in Ant Colony Optimisation", In: M. Dorigo, L.M. Gambardella, M. Birattari, M. Martinoli, R. Poli, T. Stutzle (eds.), *ANTS 2006*, LNCS 4150, pp. 484-491, Springer-Verlag Berlin/Heidelberg, 2006.

[5] L. Zadeh, "Fuzzy algorithms", *Information and Control*, Vol. 12, pp. 99-102, 1968.

[6] L. Zadeh, "Fuzzy sets", *Information and Control*, Vol. 8, pp. 338-353, 1965.

[7] L. Zadeh, "Maximizing Sets and Fuzzy Markoff Algorithms", *IEEE Transactions on Systems, Man, And Cybernetics - Part C: Applications and Review*, Vol. 28, No. 1, pp. 9-15, 1998.

[8] http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95 (TSPLIB), cited on December 15, 2007.

Gloria Cerasela Crişan, Elena Nechita
University of Bacău
Department of Mathematics and Informatics
8, Spiru Haret Str., 600114 Bacău, Romania
E-mail: {ceraselacrisan, enechita}@ub.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 232-237

Evaluation of Adaptive Radio Techniques for the under-11GHz Broadband Wireless Access

Nicolae Crişan, Ligia Chira Cremene, Emanuel Puşchiţă, Tudor Palade

Abstract: The evaluation is performed in the attempt to improve link availability and transmission quality of BWA (Broadband Wireless Access) systems, and to achieve an efficient use of radio resources. We have analysed and compared the physical and link layer performance for two common standards: the IEEE 802.16 WMAN-SC and the IEEE 802.11 (a, b, g) WLAN. The paper contains a performance comparison of 802.16 and 802.11g in terms of (a) the effect of the number of nodes and of the packet length, (b) the influence of adaptive modulation, and (c) the influence of mobility. The limitations of the 802.11a PHY are also extracted from simulations, and the BER performance of the 802.16 FEC block is simulated and discussed. The implementations, simulations and evaluation have been performed using the ns-2 simulator, Matlab-Simulink and SystemView. The detailed observations will enable us to set the premises for new adaptive techniques and strategies.
Keywords: Adaptive radio techniques, BWA (Broadband Wireless Access), FEC, Eb/No estimation, Wireless standards coexistence.

1 Introduction

As broadband wireless communications have gained increased interest during the last few years, the convergence of broadband wireless access is the next challenge in this field [2].

In this paper we first provide an overview of the main specifications of the 802.11 (WiFi) and 802.16 (WiMAX) standards [1], [3]. Then, we describe a comparative performance analysis of the wireless access segment for these two technologies. The conclusion of our analysis recommends the implementation of the IEEE 802.16 standard even for the "last-mile" wireless access system.

Nowadays BWA (Broadband Wireless Access) systems face a series of challenges, among which are: link availability, link reliability, coverage, and capacity. After identifying BWA challenges and current technical solutions, we have established our focus on physical and link layer adaptation methods. At the physical layer we can have adaptation of the processing gain, of equalization, of bandwidth and power levels, of modulation schemes or we can have adaptive antennas. At the link layer we can adapt the frame length for instance or the error correction mechanism. This article focuses on several aspects: the BER performance of the 802.16 FEC block and a performance comparison between 802.16 and 802.11 in terms of (1) the effect of the number of nodes and of the packet length, (2) the influence of adaptive modulation, and (3) the influence of mobility.

The single carrier (SCa) specifications, according to the 802.16a version [4], are designed for operation in rural or suburban areas. The most frequent envision is that of a combination of 802.16, as backhaul, and 802.11 (a, b, g, n) creating a complete, integrated, wireless solution for delivering high-speed, broadband applications to businesses, industrial and residential area. Table I (fig. 1) contains a brief comparison of the 802.11a and 802.16a specifications [4], [5], [14].

	802.11a	**802.16a**
Frequency band	5GHz	Below 11 GHz
Range cellular radius	300 feet, LOS, NLOS	30 miles, NLOS
Bit rate	Max 54 Mbps (20 MHz)	Max 75 Mbps (20 MHz)
Channels	20 MHz, 12 non-overlapping	Scalable, flexible channel width, 1.75 MHz to 20 MHz
Modulation	BPSK, QPSK, 16QAM, 64QAM	BPSK, QPSK, 16QAM, 64QAM, 256QAM
PHY format	64 FFT OFDM	Single Carrier, Adaptive 256 FFT OFDM 2048 FFT OFDMA
Coding	Convolutional ½ - ¾	FEC – RS, CC ½ - 7/8 (BTC, CTC)

Figure 1: Table I: 802.11a vs. 802.16a

Section 2 briefly discusses the benefits of existing adaptive techniques emphasizing the main improvements that these techniques bring to wireless systems operation. It also covers the main issues regarding the IEEE 802.11 standards and provides a short performance analysis of the 802.11a PHY. Section 3 analyses a performance comparison between 802.16 and 802.11 in terms of (1) the effect of the number of nodes and of the packet length, (2) the influence of adaptive modulation, and (3) the influence of mobility. The next focus area is the 802.16 FEC block, for which section 4 describes a performance analysis (for the under-11GHz 802.16 SC specifications). Our detailed conclusions are presented in section 5.

2 Benefits of Adaptive Radio Techniques

Part of our study was to identify the existing adaptive techniques and the improvements brought by each of them. We have summarized this attempt in table II (fig. 2) which contains a synthetic view of some adaptive techniques used nowadays in broadband multicarrier wireless systems, together with the benefits they bring [6].

Adaptive Techniques	**Improvements**
Adaptive modulation	12-16dB SNR increase
Adaptive power control	Efficient use of radio resources
Adaptive subcarrier allocation	Link availability and robustness through allocation of a lower modulation scheme for the low SNR subcarriers
Adaptive bandwidth	QoS increase, BER decrease A 10 times smaller bandwidth is needed, and this results in a received SNR increase of 10 dB
Adaptive user allocation	Average signal power increase of 3-5dB Bite rate and SNR increase, BER decrease Higher capacity achieved by exploitation of frequency dispersive fading, by allocating the peaks in frequency response

Figure 2: Table II: Benefits of using adaptive techniques

We have performed a set of simulations on the 802.11a PHY Simulink platform [7], aiming to establish the limitations of this scheme in terms of received SNR, bit rate, packet error rate, and number of propagation paths. The main blocks were: a data source, a modulator bank, the OFDM transmit assembly, the channel, the OFDM reception assembly, a frequency domain equalizer, and the demodulator bank. The code employs a link adaptation scheme wherein we select the best coding rate and modulation scheme based on channel conditions.

We have simulated both LOS and NLOS propagation by combining and adapting channel blocks available in the Simulink Communications Blockset: e.g. Rician Fading Channel and Multipath Rayleigh Fading Channel. We have varied the number of paths in the case of the Rayleigh channel, and modelled indoor and outdoor environments by changing the delay spread according to the standard specifications [5].

The main limitations of the 802.11a PHY, were highlighted during our Matlab-Simulink tests [15]. In order to cope with the multipath environment (indoor and outdoor) the system needs special solutions (e.g. spatial diversity). Incorporating any kind of spatial diversity scheme into an 802.11a system will obtain a high-rate packet transmission system suitable for high throughput applications like videoconferencing and multimedia [8], [16].

3 IEEE 802.11 Standard Issues and PHY Performance Evaluation

We have compared the performance of 802.16 and 802.11 in terms of (1) the effect of the number of nodes and of the packet length, (2) the influence of adaptive modulation, and (3) the influence of mobility.

For this study, we used the ns-2 simulator, version 2.29. We patched a mobility package which implements the IEEE 802.16 standard, from the National Institute of Standards and Technology, USA [9]. This patch contains an 802.16 model and a mobility package. The patch was developed in order to make modules publicly available for several wireless and wired technologies like: IEEE 802.3 (Ethernet), IEEE 802.15.1-WPAN, IEEE 802.11b-WLAN, IEEE 802.16-WMAN, and UMTS -WWAN.

We have tested the effect of the number of nodes and of the packet length for both 802.16 and 802.11 standards. We have analysed the effect of transmitting nodes (10, 50, 100 nodes) cumulative with the packet size (256, 512, 1024 bytes) for 802.16 and 802.11 scenarios (tables from fig. 3).

TABLE III
SIMULATION RESULTS FOR 100 TRANSMITTING NODES

	802.16	802.11				
Packet size [bytes]	256	512	1024	256	512	1024
Ae2ed [ms]	1,13	1,23	1,38	60	66	76
Throughput [kbps]	1400	2700	5600	14	62	240
Jitter [s]	$2x10^{-4}$	$2x10^{-4}$	$1.7x10^{-4}$	55	57	63
Tx nodes [number]	62	62	62	62	2	6

TABLE IV
SIMULATION RESULTS FOR 100 TRANSMITTING NODES AND DIFFERENT NUMBER OF PACKETS/SECOND EMITTED

	802.16					802.11
Modulation	QPSK 1/2	QPSK 3/4	16QAM 1/2	16QAM 3/4	64QAM 3/4	BPSK 1/2
Ae2ed [ms]	48	35	28	17	1,7	35
Throughput [kbps]	860	1300	1720	2820	2850	1300
Jitter [s]	45	33	24	14	0.7	27

TABLE V
CUMULATIVE EFFECT OF MOBILE NODES' SPEED AND HANDOVER DELAY

	802.16			
Modulation	1 m/s	5 m/s	15 m/s	30 m/s
Ae2ed [ms]	350	350	350	350
Handover delay [s]	0,005s	0,005s	0,005s	0,01s
Packet loss [packets]	5	8	8	21

TABLE VI
END-TO-END DELAYS FOR USER MOBILITY ON DIFFERENT WIRELESS TECHNOLOGIES

Network type	Fixed network		Wireless network
Technology type	IP core		WLAN IP core
Handover decision	Layer 3		Layer 2
Protocol	IPv4	IPv6	IEEE 802.11
Handover delay [s]	0.007s	0.006s	0.004s

Figure 3: Simulaton results

We can observe a decrease in the quality of the network parameters for 802.11 compared to 802.16. Higher the number of ready-to-transmit stations, lower the number of actual transmitting nodes. This is the result of (1) the possible reach of maximum transmitting stations in the given scenario, and (2) the medium access technique. Since 802.11 uses contention-based access mechanism (CSMA/CA), for 802.16 is used a contention-less access mechanism (TDD). Using a more appropriate access technique, for 802.16 scenarios the best results are obtained even for the average-and-to-end (Ae2ed) delay, throughput, and jitter.

In order to test the influence of adaptive modulation, we set up a scenario containing 20 nodes. For the 802.16 scenarios, we set up QPSK1/2, QPSK 3/4, 16QAM 1/2, 16QAM 3/4, 64QAM 3/4 modulations, and for 802.11 we set up BPSK 1/2. The packet size was 1024 bytes and the source generated 40 packets/second.

The IEEE 802.16 standard specifies adaptive modulation at the physical layer. Adaptive modulation enables an 802.16 system to optimize the throughput based on the propagation conditions. Using an adaptive modulation scheme, the system can choose the highest order modulation provided.

We can observe (see fig. 3) that when we use a higher order modulation, the average received throughput is higher. This scenario highlights a very important characteristic of 802.16, the adaptive modulation.

For the next generation technologies, mobility is more than a necessity, it is a requirement. New developed architectures dedicate a special attention to this aspect. In this scenario we have tested the mobility effect in a handover process. A mobile node communicates with a base station and, at a given time, the communication route is changed from one base station to another. The cell radius is 1000 m. The traffic type is CBR with 10 packets generated per second. The packet length is 4960 bytes. We vary the speed of the mobile station from 1m/s to 5m/s, 15m/s, 30m/s (table V).

Depending on which layer the roaming occurs we could define two major types of roaming: layer 2 roaming and layer 3 roaming. In order to compare 802.16 performances, handover processes were simulated also for mobile IPv4, mobile IPv6, and 802.11 [10]. Results presented in table V and table VI indicate better performances for the handover delay in the case of 802.16. A layer 2 (802.11) handover decision implies less computational time on the mobile node vs. a layer 3 (IPv4, IPv6) handover decision.

4 FEC Performance Analysis for the under-11GHz 802.16 SC Specifications

The 802.16 PHY layer includes a FEC block with two concatenated codes: Reed Solomon (RS) as an outer coder and a convolutional one (CC) as inner coder. We have implemented the IEEE 802.16 SC transmit processing chain according to the WirelessMAN-SCa PHY specifications [4], except for the burst framing and power control blocks which were not considered.

The 802.16 FEC is based on concatenated RS outer coder and rate compatible TCM inner coder, RS(255,239,8)

and TCM (K=7 rate 1). This paper analyses the 802.16a SCa PHY chain, which is modelled using SystemView. In addition, we have studied the effect of using an interleaver between the inner and the outer coder. The interleaving process is optional [4] because of the cost of the process itself, which is known to be responsible for significant group delays. Two cases have been analyzed: with and without the interleaving process. The results are evaluated considering BER as a function of Eb/No (Energy per bit per Noise).

The most important parameter is the BER and the most difficult task is to accurately estimate the Eb/No. SystemView provides different tokens that perform most of the operations necessary to implement and simulate a QPSK modulator/demodulator [11]. The Eb/No estimation method is explained in [12]. In order to find the right value for the noise power density we used a SystemView metasystem [13].

Figure 4, left side, presents the BER curve as a function of the Eb/No. The BER plot is obtained by using five different burst samples, each burst having a different Eb/No value (the initial value for the Eb/No is 3.8 dB) and a step of 0.4 dB. Plot 1 from Fig. 1 is the theoretical coherent QPSK BER as a function of Eb/No.

Plots 2 and 3 show the BER of the IEEE 802.16a SCa system, QPSK 1 Viterbi rate, with and without the interleaving process. The difference between these two plots is around 0.4 dB, due to the interleaver buffer size, a smaller buffer being almost ineffective. The spectral efficiency can be improved by increasing the gain of the FEC itself. We consider the interleaver to be efficient when the system can cope with a disruptive burst, which heavily affects the 802.16 packets. An important task is to optimize the interleaver buffer size for maximum efficiency. For the interleaver to be efficient no more than 8 disrupted bytes should be contained in one RS block (256 bytes/block). This means the distance between two consecutive disrupted bytes should be no less than 32 bytes (256/8).

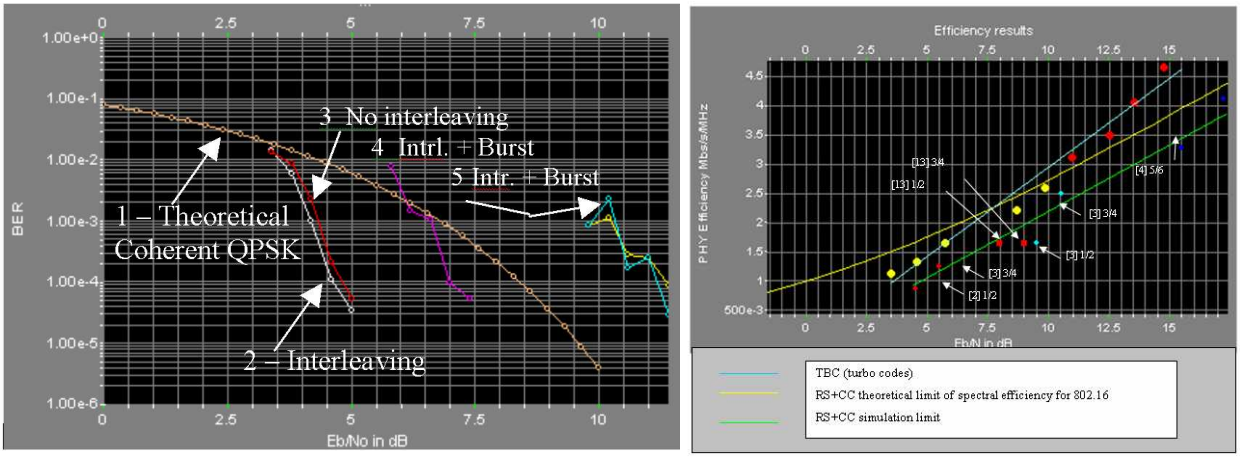


Figure 4: BER vs. Eb/No QPSK IEEE-802.16a SCa-DL (***left***) and Spectral efficiency evaluation for the 802.16 SC system (***right***)

In the presence of both Gaussian noise and the sinusoidal burst the BER degrades. The BER graphical results are shown in Fig. 1 (plots 4 and 5). Plot 1 is the well-known QPSK coherent theoretical BER. Plot 4 shows the BER curve for the 802.16a SCa QPSK 1 with the block interleaving process active and sinusoidal disruptive burst (60l's) and the same operation scheme for plot 5 but with an 80-us disruptive burst instead of a 60-l's one. The system loss due to BER degradation is about 5.5 dB. Nevertheless, without the interleaving process, the 802.16a SCa QPSK 1 system will perform worse, a number of symbols remaining uncorrected even in the case when the Eb/No ratio increases over the 11.5 dB margin.

This scenario must be considered especially when we talk about the coexistence of 802.16 with systems operating in the same band. A disruptive burst coming from such a coexistent system can affect many symbols from the 802.16 data packet. The interleaving process must perform efficiently. The ideal interleaver size depends on the length of the disruptive packets. Fortunately, current standards specify an interference avoidance mechanism based on packet splitting to decrease the length of the data packet. Using the interleaver will thus be an advantage for an 802.16 system facing interference from coexisting systems.

In figure 4, right side, the yellow line establishes the theoretical limit of spectral efficiency for 802.16.a SC system. As one can see, the blue line which points in the limit for the usage of Turbo codes goes beyond the theoretical limit for the RS+CC FEC scheme. When a disruptive burst interferes with the useful signal the PHY

efficiency remains constant as long as the FEC copes with the interference, with the cost of increasing the Eb/No ratio.

The use of the block interleaving as suggested by the 802.16 standard, appears to be inefficient if we analyze the packet delay. In the simulated case, for the maximum size of the block interleaver buffer, the delay must be 3356 Tbyte. At a 10.04 Mbps data rate and a RS (255, 239, 8), Tbyte=712ns. The delay introduced by the interleaver will be around 2.3 ms, which could be an important delay. The system robustness when facing a 70-us disruptive burst is comparable to the one of the same system using the block interleaver, but the packet time delay makes the difference.

We have tested four FEC schemes: RS+Forney, RS+Block Interleaver, TBC+Helical, and RS+BI+CC. The FEC scheme that proves to be weaker in the presence of the disruption burst is RS+BI+CC. The most robust are the RS+Forney and RS+Block Interleaver schemes which spread the bits so that the error is avoided. This means that a maximum disruptive burst of 80 l's could be handled under a heavily interfered environment with no conflict to another coexisting signal, within the same channel. As it was proved, error mitigation depends on how the FEC changes its scheme with respect to channel parameters. In [12] we have identified six evaluation criteria for an adaptive FEC scheme that would improve the performance in terms of throughput in an 802.16a system. Based on these criteria the results lead to the remark that the two concatenated RS and convolutional codes are not ready to overcome the coexistence demands in a heavily disruptive environment.

5 Conclusions

Taking into consideration that 802.11 and 802.16 technologies will coexist, complementing each other, and in order to better understand their behaviour, we need a set of parameters that are measured in both of these networks. This paper has analyzed (1) the effect of the number of nodes and of the packet length, (2) the influence of adaptive modulation, (3) the influence of mobility for both standards, and (4) the BER performance of the 802.16 FEC block. Also, the handover procedure was analyzed regarding the effect of the mobile station speed over the link parameters. The conclusion of our analysis recommends the implementation of the IEEE 802.16 standard even for the "last-mile" wireless access system.

A wireless system like the IEEE 802.16 will have to provide better quality by means of a new vision about a smart FEC block. In order to improve the PHY spectrum efficiency the FEC scheme should be based on simple and efficient techniques (lower complexity and higher applicability). Our simulations show that the smaller interleaver buffer is almost ineffective while the larger one adds important delays. From this point of view the usage of the block interleaving is not suited for the 802.16a system. The convolutional interleaving scheme performs better from the delays point of view diminishing the delay from 2.3 ms to 67 us. Taking these into account it is obvious that the use of the convolutional Forney interleaver instead of the block interleaver could be a better choice in order to decrease the packet delay. In a heavily disturbed environment, an adaptive FEC scheme, based on the above criteria, could improve the BER performance and cope with the interferences.

References

[1] IEEE 802.16 *Standard - Local and Metropolitan Area Networks* – Part 16, IEEE Std 802.16a-2003.

[2] WiMAX Forum, *Fixed, Nomadic, Portable and Mobile Applications for 802.16-2004 and 802.16e WiMAX Networks*, Nov. 2005.

[3] IEEE standard for local and metropolitan area networks part 11: wireless LAN MAC and PHY specifications, 1999.

[4] IEEE Standard for Local and metropolitan area networks – 802.16a Part 16: Air Interface for Fixed Broadband Wireless Systems – Amendment 2: MAC Modifications and Additional PHY Specifications for 2-11 GHz, IEEE– 1 April 2003.

[5] IEEE Std 802.11a-1999 (Supplement) Telecommunications and information exchange between systems. Local and Metropolitan Area Networks. Specific requirements Part 11: Wireless LAN MAC and PHY specifications High-speed Physical Layer in the 5 GHz Band.

[6] Eric Phillip Lawrey, *Adaptive Techniques for Multiuser OFDM*, Ph.D thesis, James Cook University, December 2001.

[7] Martin Clark, *MATLAB Central model: IEEE 802.11a WLAN PHY*, 2003.

[8] Mohinder Jankiraman, *Space-time Codes and MIMO Systems*, Artech House, 2004.

[9] National Institute of Standards and Technology, USA, http://www.antd.nist.gov/seamlessandsecure/toolsuite.html

[10] E. Puşchiţă, T. Palade, F. Sandu, *Intra-System Mobility Evaluation for Different Wireless Technologies*, 2006 International Conference on Wireless and Mobile Communications – ICWMC 2006, IEEE Computer Society, Bucureşti, RO, July 29–31, 2006.

[11] A. Ganz, Z. Ganz, K. Wongthavarawat, *Multimedia Wireless Networks: Technologies, Standards, and QoS*, Prentice Hall PTR, 2003.

[12] N. Crisan, L. Chira Cremene, E. Puschita, *FEC Performance Analysis for the under-11GHz 802.16 SC Specifications*, 16th IST Mobile and Wireless Communications Summit, Budapest, Hungary, ISBN 978-963-8111-66-1, July, 1-5, 2007.

[13] I. Koffman and V. Roman, *Broadband wireless access solutions based on OFDM access in IEEE 802.16,* IEEE Communications Magazine, vol. 40, no. 4, pp. 96-103, Apr. 2004.

[14] Ekram Hossain, *IEEE802.16/WiMAX-Based Broadband Wireless Networks: Protocol Engineering, Applications, and Services*, IEEE Communication Networks and Services Research, 2007.

[15] Ligia Chira, Tudor Palade, Emanuel Puschita, *Performance analysis of spatial diversity schemes on an 802.11a PHY platform*, 7th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Telsiks'05, vol.1, p.101-104, ISBN 0-7803-9164-0, catalog no. 05EX1072, Nis, Serbia, 28-30 September 2005.

[16] L. Chira Cremene, T. Palade, *The Adaptive Potential of Space Diversity Techniques*, The Mediterranean Journal of Electronics and Communications, SoftMotor Ltd. ISSN: 1744-2400, Vol. 3, No. 3, pp.100-109, 2007.

Nicolae Crişan, Ligia Chira Cremene, Emanuel Puşchiţă and Tudor Palade
Technical University of Cluj-Napoca
Communications Department
400027, 26-28 Baritiu street, Romania
E-mail: nicolae.crisan@com.utcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 238-242

A First Derivate Based Algorithm for Anomaly Detection

Petar Čisar, Sanja Maravić Čisar

Abstract: In attack continuum three different periods can be recognized - before, during and after the exploitations of a vulnerability. This paper is related to the second period and will focus on the initial start of an attack and that type of attacks that are detectable in real time. A lot of different approaches exists for anomaly detection. One of them is behavioral analysis, thus in accordance with this, a threshold algorithm based on first derivate is presented.
Keywords: attack, anomaly detection, first derivate

1 Introduction

During normal traffic, there are lot of increases and decreases of traffic volume. The algorithm formulated here differentiates between slopes of two volume increases: the increase of normal traffic and the increase of traffic volume at the beginning of attack, irrespectively of the actual volume or time. It relies on the author's standpoint that at that short starting time interval, the increase of traffic is sharper than it is in any other case of normal traffic.

2 Characteristics of Network Intrusions

At the very beginning of an attack, aggregate traffic volume (trace + attack) rapidly increases with sharp slope over a short time period ([7], [9] - Figures 1, 2 and 3). The amplitude of this increase is proportional to the intensity of attack. According to [7], in the case of a small attack the mean amplitude of attack could be approximated with 50% of the traffic's actual mean, while for intensive attack this mean amplitude is 250% of the traffic's mean.

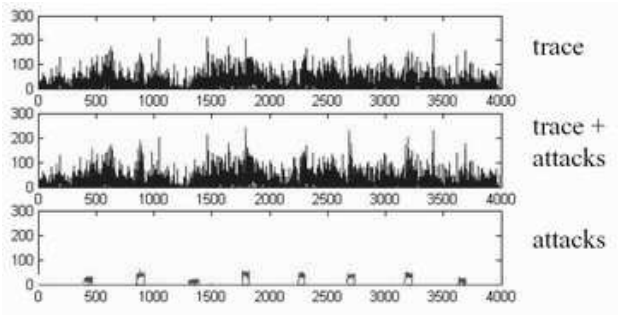


Figure 1: Small attack

The detection of small attacks is important for several reasons. Firstly, early detection of attacks with increasing intensity would enable defensive actions to be taken earlier. Secondly, detection of small attacks would enable the detection of attacks close to the sources, since such a placement of detection algorithms can facilitate the identification of stations that are participating in a distributed denial of service (DoS) attack.

Let us assume that for a given network traffic in some time interval, each connection is assigned a score value, represented as a vertical line (Figure 3. - [9]). The dashed line in Figure 3. represents the real attack curve that is 0 for non-intrusive (normal) network connections and 1 for intrusive connections. Response time represents the time elapsed from the beginning of the attack till the moment when the first network connection has the score value higher than prespecified threshold.

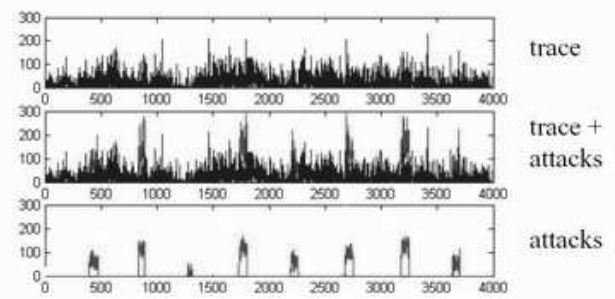


Figure 2: Intense attack

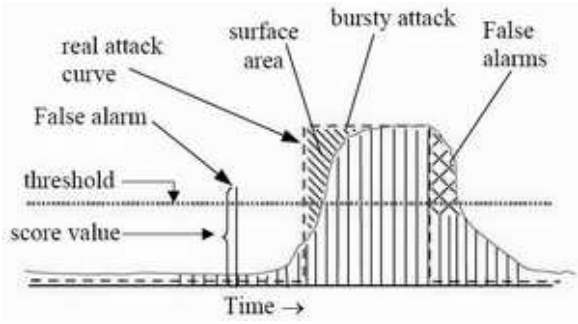


Figure 3: Attack curve

3 A first Derivate Based algorithm

This algorithm is based on the behavioral analysis of network traffic. The beginning of an attack is characterized by short-term amplitude increase.

Using some network monitoring software (NMS) during the period of normal traffic it is possible to determine the maximal value of the traffic curve's first derivate which is considered normal. Every consecutive positive deviation from that fixed threshold is attack suspicious event and should be detected by NMS.

The definition of the first derivate gives starting platform for further calculations:

$$a'(t) = \lim_{t \to 0}(\frac{\Delta a}{\Delta t}) \tag{1}$$

Namely, the first derivate in some point, according to its definition, determines the slope of the curve (more precisely, the tangent of the curve) in that point.

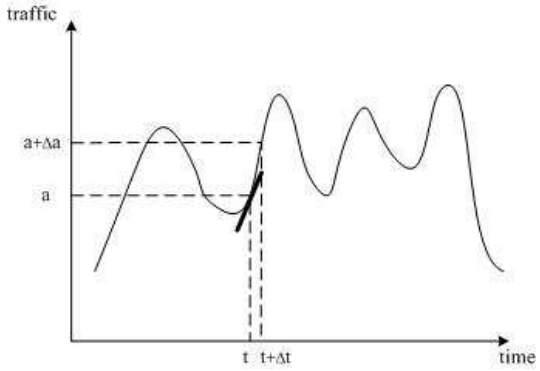


Figure 4: Definition of the first derivate

For a small enough Δt, the following approximative relation will be used:

$$a'(t) \approx (\frac{\Delta a}{\Delta t}) \tag{2}$$

The maximal slope in normal traffic is determined as the maximal value of all calculated first derivates in the observation interval T (Table 1.) and represents the value of fixed threshold:

$$\Delta a_{max} = max(\Delta a_i), \text{where } \Delta a_i = a_i - a_{i-1},\ i = 1,2,3,\ldots \tag{3}$$

t	0	Δt	$2\Delta t$	$3\Delta t$	$4\Delta t$	$5\Delta t$	...
a	a_0	a_1	a_2	a_3	a_4	a_5	...
	Δa	Δa_1	Δa_2	Δa_3	Δa_4	Δa_5	...

Table 1: Determination of maximal slope

The approximation of the first derivate is more precise as Δt is closer to zero. Besides, it is important to establish such small Δt, with which NMS is able to operate.

To allow the NMS enoughly time to receive and analyze an accurate representation of normal traffic, the authors of this paper recommend to let the initial training period run for at least 4 days (2 working and 2 weekend days) before terminating this phase.

4 Real-time Detection

Considering the variety of attacks, it is rather difficult to precisely define the starting part of the attack timeline. Some attacks are immediately recognizable, perhaps taking the form of one or more packets operating over a short time period Δt - e.g. less than one second [10]. Others are active for a much longer period (e.g. hours, days or even weeks) and may not even be identified as attacks until a vast collection of event records are considered in aggregate. Thus, while every attack has a definite beginning, this starting point is not always discernible at the time of occurence.

The main idea of this paper is the detection of such type of attacks that are recognizable in real time (real-time detection) - true time zero plus some arbitrary small time slice beyond that -i.e. less than a few seconds [10]. Real-time, according to industry definitions, can be expressed like time interval 5 s - 5 min.

Based on what has been stated in Section 3 and 4, the authors are of the opinion that the interval $\Delta t = 1s$ satisfies all the requirements and could be accepted as adequate.

From definition of the maximal first derivate, Figure 4 and calculated Δa_{max} for normal traffic, the following relation is of significance:

$$a'_{max(t)} = \tanh\ \alpha_{max} = \frac{\Delta a_{max}}{\Delta t},\ \text{where } \Delta t = 1 \tag{4}$$

Figure 5 represents a situation during an attack:

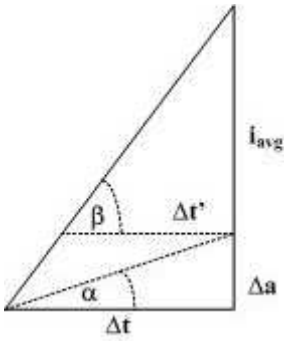


Figure 5: Attack analysis

$$\tanh \beta = \frac{i_{avg}}{\Delta t'} \tag{5}$$

where i_{avg} is the average of intrusion (attack), while $\Delta t'$ is the time during which the attack exceeds the value of i_{avg} - i.e. time zero.

As $\beta \geq \alpha$, then is $\tanh \beta \geq \tanh \alpha$. After the substitution of appropriate values the result is:

$$\Delta t' \leq \Delta t \cdot \frac{i_{avg}}{\Delta a} \tag{6}$$

The approximated value of time zero is minimal in case of a small attack. Then is $i_{avg} \approx 0.5 \cdot a_{avg}$ and $\Delta a = \Delta a_{max}$. Thus, we arrive at:

$$\Delta t' \leq \Delta t \cdot \frac{0.5 \cdot a_{avg}}{\Delta a_{max}} \tag{7}$$

5 Optimization of the Algorithm

Section 3 explained the algorithm for determining the maximum of first derivate, for accepted starting time-step Δt. But, it is not the same if accepted time-step is Δt, $2\Delta t$ or $3\Delta t$. If the time-step is longer, the computer resources work with smaller number of samples, used databases are smaller and faster for searching, among others.

On the basis of the completed first measurement with the time-step Δt (which gave Δa_{max} as maximal first derivate), by appropriate extracting of the first derivates, it is possible to create new arrays with the time-steps $2\Delta t$, $3\Delta t$ and so on (Table 2.). With an acceptible small difference between derivates $\varepsilon \to 0$ (e.g. $\varepsilon = 10^2$), searching for derivates close enough to Δa_{max} in different subarrays should be performed. In accordance with the above said, it is optimal to accept the highest of the found values. For instance:

time-step=Δt	...	Δa_{max}	...
time-step=$2\Delta t$	...	$\Delta(a_{max})_2$	...

Table 2: Optimization

where $\varepsilon \leq \mid \Delta(a_{max})_2 - \Delta a_{max} \mid$. In this case the time-step $2\Delta t$ could be accepted as optimal, which implicate that measurement intervals should be 2 s in length (0, 2 s, 4 s, 6 s, ...).

6 Acceleration of searching

The process of searching the maximal first derivate mentioned in Section 3 can be accelerated in the following way:

- we accept the values for $\Delta a_{max} = 0$ as initial maximum and ε
- then calculate Δa_1
- if $\mid \Delta a_1 - \Delta a_{max} \mid \geq \varepsilon$, then $\Delta a_{max} = \Delta a_1$ - if not, Δa_{max} stays maximum
- then calculate Δa_2
- if $\mid \Delta a_2 - \Delta a_{max} \mid \geq \varepsilon$, then $\Delta a_{max} = \Delta a_2$ - if not, Δa_{max} stays maximum
- ...
- if $\mid \Delta a_i - \Delta a_{max} \mid \geq \varepsilon$, then $\Delta a_{max} = \Delta a_i$ - if not, Δa_{max} stays maximum

In this way, by the process of dynamic update of the maximum first derivate, searching can be automatized, giving the result immediately after the end of training period and eliminating subsequent data analysis.

7 Summary and Conclusions

The algorithm presented in this paper is relatively simple and easily applicable in NMS systems. It has good sides and bad sides, as well. The advantage is that it is based on an algorithm which is easy to realize by NMS, using built functions for triggering and just one single threshold. Also, due to the nature of the fixed threshold, this algorithm is insensitive to extremly slow attacks. Its disadvantage is the relatively long, but simple training period.

References

[1] S. Sorensen, "Competitive Overview of Statistical Anomaly Detection," *White Paper, Juniper Networks*, 2004.

[2] G. Fengmin, "Deciphering Detection Techniques: Part II Anomaly-Based Intrusion Detection," *White Paper, McAfee Security*, 2003.

[3] "SANS Intrusion Detection FAQ: Can you explain traffic analysis and anomaly detection?", *http://www.sans.org/resources/idfaq/anomaly_detection.php*

[4] "Intrusion detection systemWikipedia," *http://en.wikipedia.org/wiki/Intrusion-detection_system*

[5] M. Douglas, "Introduction to Statistical Quality Control," *5th Edition, John Wiley & Sons*, 2005.

[6] Statistical Quality Control, *www.wiley.com/college/reid/0471347248/samplechapter/ch06.pdf*

[7] S. A. Vasilios, "Denial of Service and Anomaly Detection," *SCAMPI BoF, Zagreb*, 2002.

[8] CAIDA, the Cooperative Association for Internet Data Analysis: Inferring Internet Denial of Service Activity, *University of California, San Diego*, 2001.

[9] A. Lazarevic, E. Levent, O. Aysel, K. Vipin, S. Jaideep, "A comparative study of anomaly detection schemes in network intrusion detection,"
http://www-users.cs.umn.edu/∼aleks/MINDS/talks/siam03.pdf

[10] M. Roesch, "Next-generation intrusion prevention: Time zero (during the attack),"
http://searchsecurity.techtarget.com/tip

Petar Čisar
Telekom Srbija
Prvomajska 2-4, Subotica, Serbia
E-mail: petarc@telekom.yu
Sanja Maravić Čisar
Polytechnical Engineering College
Marka Oreškovića 16, Subotica, Serbia
E-mail: sanjam@vts.su.ac.yu

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 243-247

COMDEVALCO Development Tools for Procedural Paradigm

István Gergely Czibula, Codruţ-Lucian Lazăr, Ioan Lazăr, Simona Motogna, Bazil Pârv

Abstract: This paper presents an Agile MDA (*Model-Driven Architecture*) process, based on test-driven development methodology, and a Component Definition, Validation, and Composition (COMDEVALCO) *Workbench* that allows developers to build, run and test executable models in short, incremental, iterative cycles. The system behavior is modeled with UML Structured Activities, which give a complete and precise description. Testing and debugging techniques implemented comply to OMG Model-level Testing and Debugging Specification and UML Testing Profile.
Keywords: Computer-aided software engineering, Object-oriented design methods, Debugging aids, Testing tools.

1 Introduction

MDA framework [8] refers to the specification of software systems in a platform-independent way (PIM - Platform-Independent Models) and their transformation into a platform-specific one (PSM - Platform-Specific Models). MDA emphasizes the idea of using models (usually UML diagrams) as primary artifacts throughout the software engineering lifecycle. Because development processes based on MDA are heavy-weight, *agile* software development methodologies were developed, applying agile principles and defining methods for delivering small slices of working code as soon as possible: models have to be constructed, run, tested and modified in short incremental, iterative cycles.

UML 2 is the de-facto standard for modeling software systems, but agile processes tend to minimize the usage of UML models. To be ready for applying agile principles, models must act just like code [4]. A model is *executable* if it contains a complete and precise behavior description. Because UML does not cover model execution, UML 2 Action Semantics [11] were introduced in order to provide the necessary foundation. But building such an UML-based model is a tedious task or even an impossible one, because of many UML semantic variation points.

This paper is part of a series [12, 2, 13] referring to COMDEVALCO - a framework for definition, validation, and composition of software components. Its main constituents are: an object-based modeling language, a toolset, and a component repository. Paper describes the model definition and validation tools, included in the COMDEVALCO *Workbench*.

The paper is organized as follows: after this introductory section, the second discusses related work and sketches our proposal. Next section describes the COMDEVALCO *Workbench*, mainly its interface related to modeling, debug, and model simulation. Fourth section presents an example application developed using the workbench, while the last one contains some conclusions and outlines the future work.

2 Agile MDA and executable models

2.1 Related work

Some independent efforts aiming to develop environments for manipulating executable models are in progress. OMEGA Toolset, Pathfinder, and Kermeta are relevant examples.

OMEGA Toolset [7] provides development based on formal techniques for embedded and real-time systems. The development methodology is based on a subset of the UML modeling and specification capabilities, including: class diagrams and state machines, Object Constraint Language (OCL), use case diagrams, and an extension of UML sequence diagrams. PathFinder [14] offers model-based development solutions for systems engineering and software development. It facilitates the construction of PIMs, then their translation to PSMs, i.e. executable systems. Kermeta [5] is a metaprogramming environment based on an object-oriented DSL (Domain Specific Language) optimized for metamodel engineering. Kermeta allows model and meta-model prototyping and simulation.

The ideas behind existing proprietary tools are quite similar. The process of creating executable models is as follows: (a) system decomposition into a set of components, (b) component modeling using class diagrams, (c) modeling of class behavior using state machines, and (d) specification of the actions used in state diagrams using a proprietary action language.

Two categories of *action languages* were developed so far: [1, 15, 16], resembling the basic concepts of UML 2 Action Semantics, and [6], extending OCL query expressions and adding side-effect capabilities to OCL. Unfortunately, all these languages provide only a concrete syntax and does not provide graphical notations for activity diagrams. Also, the model execution needs an additional step: transforming PIM into a PSM. This way, the time delay between model change and model execution prevents rapid prototyping.

2.2 Our Solution

The proposed solution, COMDEVALCO *Workbench*, helps developer to build, test, and execute models based on UML structured activities.

The concrete syntax and graphical notations used for UML structured activities are defined in PAL Action Language [2]. PAL allows both textual and graphical notations, thus simplifying the construction of UML structured activities. Also, the workbench implements testing and debugging techniques which comply to the Model-level Testing and Debugging Specification [10] and the UML Testing Profile [9].

3 COMDEVALCO Development Tool

COMDEVALCO *Workbench* is an Eclipse RCP [3] application having three main components: *PAL Graphical Editor* (builds models using *PAL* graphical noations), *PAL Textual Editor* (builds models based on *PAL* textual notations) and *Model Executor* (an engine for executing, testing and debugging models).

The internal representation of the modeled artifacts uses the proposed object model (Figure 1): the test-cases, operations, statements, pre- and post-conditions are stored as instances of the classes in the model. The resulting *object model* is updated using both editors (the changes made in one editor being immediately visible in the other) and is executed and tested with the *ModelExecutor*.

The main classes in the diagram are: *TestCase* (developer-defined test), *Operation* (model for an operation), *Statement* and all its subclasses (models for structured programming statements: *assignment*, *if*, *while*, etc.).

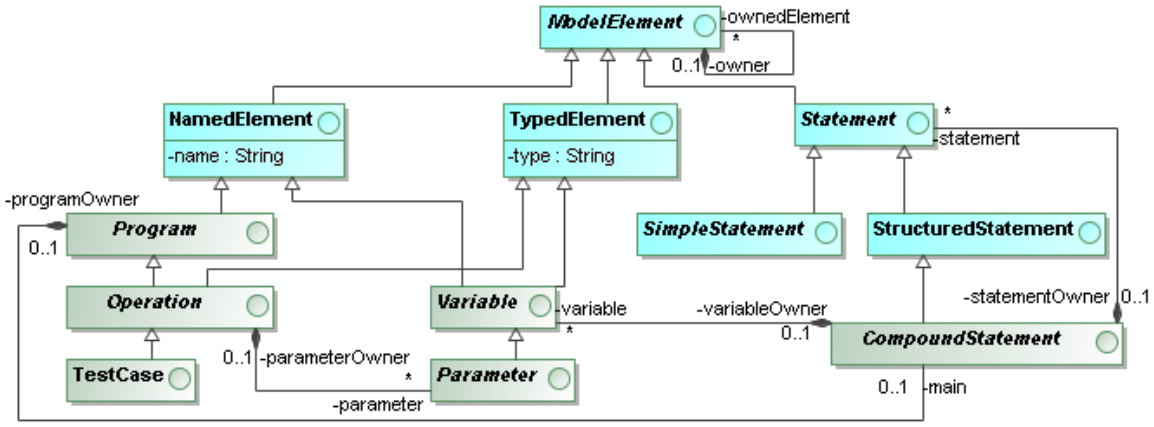


Figure 1: PAL Metamodel extract

3.1 COMDEVALCO Workbench

The model of a software system is contained in a *project*, shown in *ProjectView* (Figure 2-(1)). It includes several packages, each containing test cases and operations. The workbench interface offers a Modeling Perspective (Figure 2) and a Debug/Simulation Perspective (Figure 3).

Using the **Modeling Perspective**, the developer models the software system with the *PalTextEditor* (Figure 2-(2)) and/or the *PalGraphicalEditor* (Figure 2-(3)). This perspective has an *OutlineView* (Figure 2-(4)), displaying the structure of the operations, and a *PropertiesView* (Figure 2-(5)), showing the properties of the model elements. A *ConsoleView* and *ErrorLogView* (Figure 2-(5)) are also available; these views represent the standard input/output.

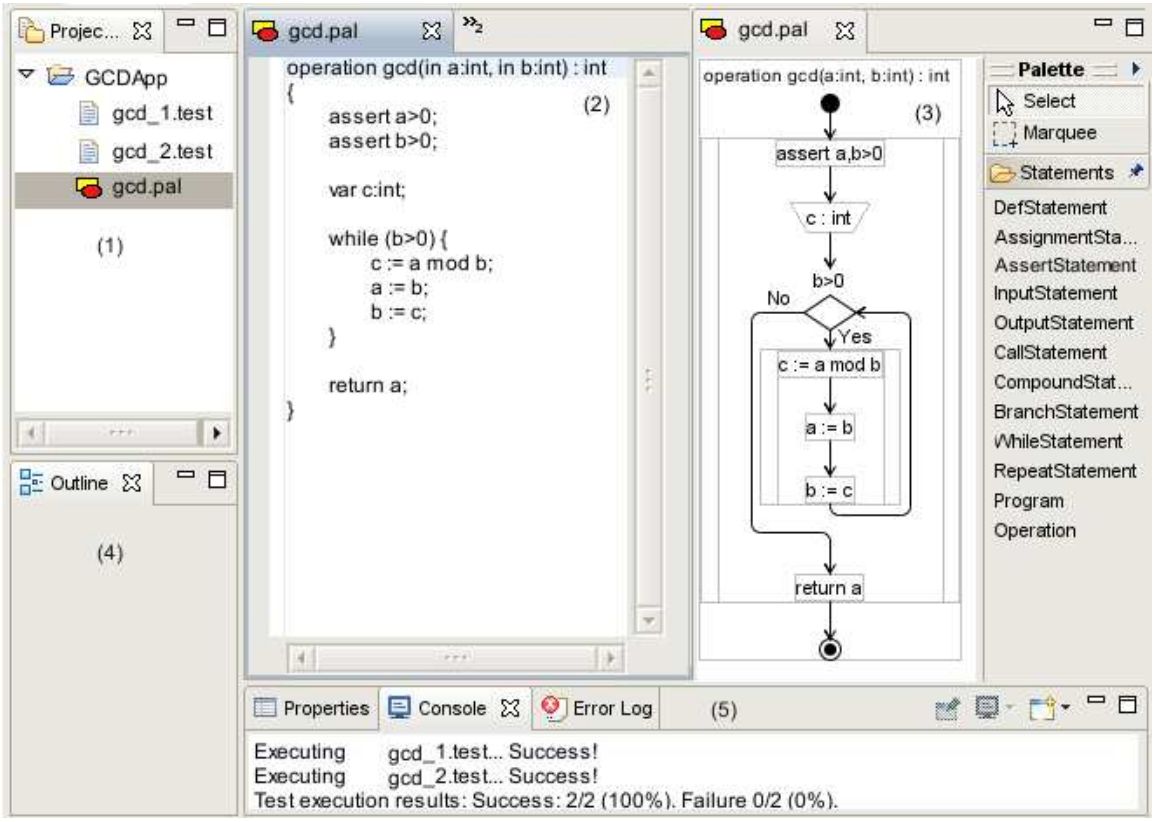


Figure 2: Modeling Perspective

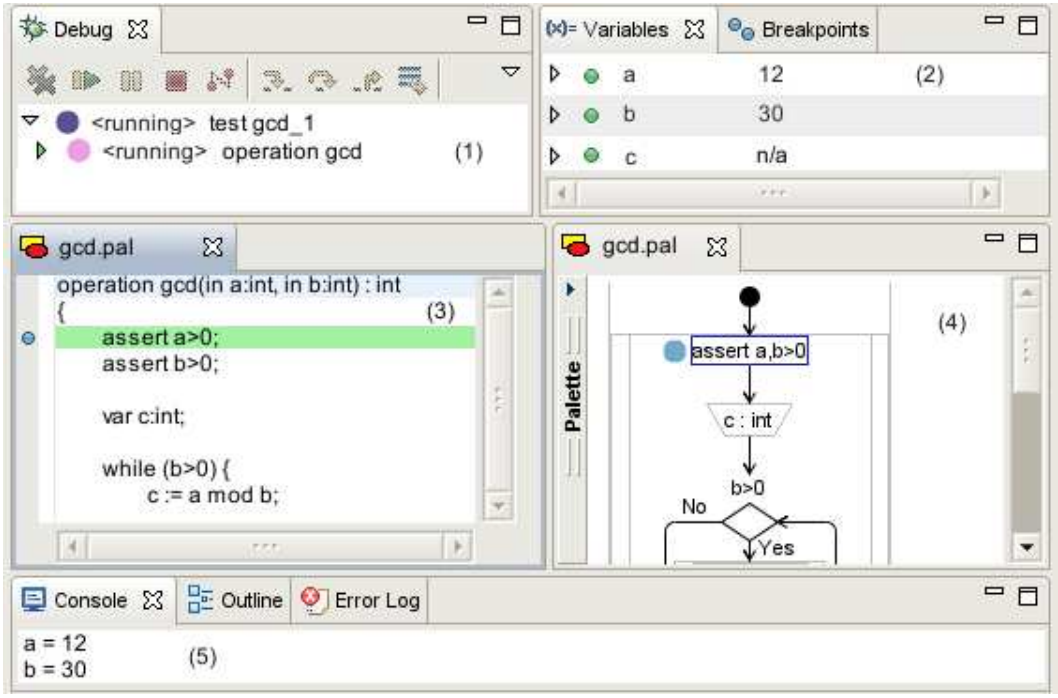


Figure 3: Debug/Simulation Perspective

The **Debug/Simulation Perspective** allows the developer to execute the model in *debug* mode. It contains a *DebugView* (Figure 3-(1)), displaying the invocation chain between operations, up to the breakpoint. Breakpoints are shown in the *BreakpointsView* (Figure 3-(2)); when a breakpoint is reached, the execution is suspended and the user can resume it or execute the program step-by-step. By clicking on an operation in the *DebugView*, its state is displayed in the *VariablesView* (Figure 3-(2)). During the debug process (Figure 3-(3,4)), both editors are available, as well as the *ConsoleView* and *ErrorLogView* (Figure 3-(5)) panels.

3.2 PalTextEditor

This editor automatically creates default implementations for attributes and referred but un-implemented operations. Following the *test-first* design approach, the developer has to create a test case, then to run the test in order be sure that the test fails, and finally to implement and update the code. During design process, the editor reports all not-implemented elements and provides an automatic way to create default implementations for them.

Based on the *PAL* source code, the editor automatically updates the object model and performs the syntactic analysis, providing an immediate feedback for the developer. Syntactic errors are reported and can be analyzed using *ErorLogView*.

3.3 PalGraphicalEditor

This editor allows the user to create programs or operations described by structured programming statements (compound statement, sequence, decision, loop) by using graphical primitives defined in PAL [2]. Other simple statements like: variable declaration, assignment, procedure call, etc., are also supported.

The editor performs automatic layout of graphical objects, so the user is not concerned with their position on the screen, being able to concentrate on the actual modeling tasks. Also, the editor supports breakpoint management and shows feedback during debug sessions.

3.4 Model Executor

COMDEVALCO *Workbench* uses a virtual machine, *Model Executor*, that executes the object model of the target system, structured as a tree (like an Abstract Syntax Tree).

The *Model Executor* walks through the tree and directly interprets each model element. This *direct interpretation* offers the opportunity to play with the model (i.e. to change the model and see the results immediately), allowing developers to run and test executable models in short, incremental, iterative cycles. Also, *Model Executor* offers debugging support for any program or operation, as it was described above.

4 Example

COMDEVALCO *Workbench* is organized in a way that encourages test-driven development. The main steps of using it are described below, using a greatest common divisor (*gcd*) sample application. First step is to create a new project for the target application. Next steps, detailed below, create tests, operations or programs.

Using the **Modeling** perspective, a new test-case is first added to the project, then constructed and executed in order to assure that the test will fail. The next step is the implementation of the functionality for which the test-case was created. For our example, the test-case invokes an operation named *gcd* inside the test, with two integer arguments (12, 30) and asserts that the result is 6 (see Figure 3). At this moment, the tool creates a model for this test-case made of a root *Testcase* element with a *CompoundStatement*, which, in turn, contains a *CallStatement* and an *AssertStatement*.

The tool generates a stub *gcd* operation with the proper list of parameters and a default body; its model is *Operation* element with two *Parameter* elements and a body - a *CompoundStatament* containing a *ReturnStatement*. The developer models this operation just like the test-case; a possible implementation is shown in Figure 2.

When the operation modeling is complete, the test-case is executed again. If the tests fail, the debug mode can be used. Figure 3 shows the execution of the test-case stopped at a breakpoint that was set in the *gdc* operation. When all tests pass, the workbench allows the restructuring of the written code, both for test-cases and operations, in order to improve the code structure.

5 Conclusions and Further Work

This paper describes a development tool, COMDEVALCO *Workbench*, allowing developers to construct, run and test executable models in short, incremental, iterative cycles. The use of PAL Language reduces the complexity of building executable models, by simplifying the construction of UML structured activities. Using a virtual machine that executes the model by direct interpretation of the model elements, COMDEVALCO *Workbench* eliminates the time delay between the model change and model execution.

In the future, we intend to extend COMDEVALCO *Workbench* in order to include object-oriented and component-based language constructs, as well as model transformation capabilities.

Acknowledgements

This work was supported by the grant ID_546, sponsored by NURC - Romanian National University Research Council (CNCSIS).

References

[1] Carter, K., *The Action Specification Language Reference Manual*, 2002. http://www.kc.com/

[2] Lazăr, I., Pârv, B., Motogna, S., Czibula, I.G., Lazăr, C.L., An Agile MDA Approach for Executable UML Activities, *Studia UBB, Informatica*, LII, No. 2, 2007, pp. 101-114.

[3] McAffer, J. and Lemieux, J.-M., *Eclipse Rich Client Platform: Designing, Coding, and Packaging Java Applications*, Addison-Wesley Professional, 2005.

[4] Mellor, S.J., *Agile MDA*, Technical Report, Project Technology, 2005. http://www.omg.org/mda/mda_files/AgileMDA.pdf

[5] Muller, P.A. et al, On executable meta-languages applied to model transformations, *Model Transformations In Practice Workshop*, Montego Bay, Jamaica, 2005.

[6] Muller, P.A., Studer, P., Fondement, F., and Bézivin, J. Platform Independent Web Application Modeling and Development with Netsilon, *SoSym*, 4, No. 4, 2005, pp. 424-442.

[7] *Omega toolset*, http://www-omega.imag.fr/tools.php

[8] Object Management Group, *MDA Guide Version 1.0.1*, 2003. http://www.omg.org/docs/omg/03-06-01.pdf

[9] Object Management Group. *UML 2.0 Testing Profile Specification*, 2005, http://www.omg.org/cgi-bin/apps/doc?formal/05-07-07.pdf.

[10] Object Management Group, *Model-level Testing and Debugging*, 2007, http://www.omg.org/cgi-bin/doc?ptc/2007-05-14/

[11] Object Management Group, *UML 2.1.1 Superstructure Specification*, 2007, http://www.omg.org/cgi-bin/doc?ptc/07-02-03/

[12] Pârv, B., Motogna, S., Lazăr, I., Czibula, I.G., Lazăr, C.L., COMDEVALCO - a framework for software component definition, validation, and composition, *Studia UBB, Informatica*, LII, No. 2, 2007, pp. 59-68.

[13] Pârv, B., Lazăr, and Motogna, S., COMDEVALCO framework - the modeling language for procedural paradigm, *IJCCC*, 3, No. 2, 2008 (in print).

[14] Pathfinder project, *Pathfinder MDA* http://www.pathfindermda.com/index.php

[15] ProjTech AL: Project Technology, Inc, *Object Action Language*, 2002.

[16] Telelogic AB, *UML 2.0 Action Semantics and Telelogic TAU/Architect and TAU/Developer Action Language*, Version 1.0, 2004.

István Gergely Czibula, Codruţ-Lucian Lazăr, Ioan Lazăr, Simona Motogna, Bazil Pârv
Babeş-Bolyai University
Faculty of Mathematics and Computer Science
Department of Computer Science
1, M. Kogălniceanu, Cluj-Napoca 400084, România
E-mail: {istvanc,ilazar,motogna,bparv}@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 248-252

Hierarchical Clustering Based Design Patterns Identification

István Gergely Czibula, Gabriela Şerban

Abstract: *Design patterns* have attracted significant attention in software engineering in the last period. An important reason behind this is that design patterns are potentially useful in both development of new, and comprehension of existing object-oriented design, especially for large legacy systems without sufficient documentation. That is why the problem of design patterns identification is very important. Automating the detection of design pattern instances could be of significant help to the process of reverse engineering large software systems. In this paper we aim at introducing a search based approach for identifying instances of design patterns in a software system. An experimental evaluation of our approach is also provided.
Keywords: Design patterns, searching, clustering.

1 Introduction

Software *design patterns* [3] are well-known and frequently reused micro-architectures: they provide proven solutions for recurring design problems in certain contexts. A pattern description encompasses its static structure, in terms of classes and objects participating to the pattern and their relationships, but also the behavioral dynamics, in terms of exchanged messages. The description of the solution tries to capture the essential insight which the pattern embodies, so that others may learn from it, and make use of it in similar situations: patterns help create a shared language for communicating insight and experience about these problems and their solutions.

From a program understanding and reverse engineering perspective, extracting information from a design or source code is very important, the complexity of this operation being essential. Localizing instances of design patterns in existing software can improve the maintainability of software. Automatic detection of design pattern instances is probably a useful aid for maintenance purposes, for quickly finding places where extensions and changes are most easily applied.

It would be useful to find instances of design patterns especially in designs where they were not used explicitly or where their use is not documented. This could improve the maintainability of software, because larger chunks could be understood as a whole.

The aim of this paper is to introduce an original clustering based approach for identifying instances of design patterns in an existing software system.

The rest of the paper is structured as follows. Section 2 presents our clustering based approach for identifying instances of a given design pattern in an existing software system, and Section 3 contains an experimental evaluation of it. Section 4 presents some existing approaches in the field of automatic design patterns identification. Conclusions of the paper and future research directions are given in Section 5.

2 A clustering based approach for design patterns identification

Let $S = \{C_1, C_2, ..., C_n\}$ be a software system, where C_i, $(1 \leq i \leq n)$, is an application class from the system. n represents the number of application classes from S.

A given *design pattern* p can be viewed as a pair $p = (\mathcal{C}_p, \mathcal{R}_p)$, where: $\mathcal{C}_p$ represents the set of classes that are components of the design pattern p; $\mathcal{R}_p$ is a set of constraints (relations) existing among the classes from $\mathcal{C}_p$, constraints that characterize the design pattern p. Consequently, each constraint $r \in \mathcal{R}_p$ is a relation defined on a subset of classes from $\mathcal{C}_p$.

We mention that all the constraints from $\mathcal{R}_p$ can be expressed as binary constraints (there are two classes involved in the constraint). That is why, in the following, we will assume, without loosing generality, that all the constraints from $\mathcal{R}_p$ are binary.

Let us denote by *cmin* the minimum number of binary constraints from $\mathcal{R}_p$ that a class from $\mathcal{C}_p$ can satisfy, as indicated in Equation (1).

$$cmin = min_{C \in \mathcal{C}_p} |\{r \in \mathcal{R}_p \mid \exists C' \in \mathcal{C}_p, C' \neq C \; s.t. \; C \; r \; C' \vee C' \; r \; C\}| \tag{1}$$

In this section we are focusing on identifying all the instances of a given design pattern p in a given design (software system). It can be easily seen that the problem of identifying all instances of the design pattern p

in the software system S is a *constraint satisfaction problem* [7], i.e., the problem of searching for all possible combinations of $|\mathcal{C}_p|$ classes from S such that all the constraints from $\mathcal{R}_p$ to be satisfied.

It is obvious that a brute force approach for solving this problem would lead to a worst case time complexity of $O(n^{|\mathcal{C}_p|})$. The main goal of the search based approach that we propose in this section in order to find all instances of a design pattern p is to reduce the time complexity of the process of solving the analyzed problem.

The main idea of our approach is to obtain a set of possible pattern candidates (by applying a preprocessing step on the set S) and then to apply a divisive clustering algorithm in order to obtain all instances of the design pattern p. Our search-based approach for identifying instances of design patterns in a software system consists of four steps: **Data collection**, **Preprocessing**, **Grouping**, and **Design pattern instances recovery**. In the following we will give a description of the above enumerated steps.

2.1 Data collection

During this step, the existing software system S is analyzed in order to extract from it the relevant entities: classes, methods, attributes and the existing relationships between them. In order to verify the constraints $\mathcal{R}_p$ of the design pattern p, we need to collect from the system information such as: all interfaces implemented by a class, the base class of each class, all methods invoked by a class, all possible concrete types for a formal parameter of a method, etc.

In order to express the dissimilarity degree between any two classes from S relating to the considered design pattern p, we will consider the distance $d(C_i,C_j)$ between two classes C_i and C_j from S given by the number of binary constraints from $\mathcal{R}_p$ that are not satisfied by classes C_i and C_j. It is obvious that as smaller the distance d between two classes is, as it is more likely that the two classes are in an instance of the design pattern p. The distance is expressed as follows.

$$d(C_i,C_j) = \begin{cases} 1+|\{r \,|r \in \mathcal{R}_p \; s.t. \; \neg(C_i \; r \; C_j \vee C_j \; r \; C_i)\}| & i \neq j \\ 0 & i = j \end{cases}. \tag{2}$$

Based on the definition of d given above it can be simply proved that d is a semimetric function.

2.2 Preprocessing

After the **Data collection** step was performed and the needed data was collected from the software system in order to compute the *distances* between the classes (Equation (2)), a preprocessing step is performed in order to reduce the search space, i.e., the set of possible pattern candidates.

In order to significantly reduce the search space, we eliminate from the set S of all classes those classes that certainly can not be part of an instance of the design pattern p. By applying this filtering, we will obtain a set of possible pattern candidates, denoted by $Cand(S)$. More specifically, the following filtering step is performed:

- We eliminate from the set of all classes those classes that satisfy less than *cmin* binary constraints from $\mathcal{R}_p$. The idea is that based on the definition of *cmin* given by Equation (1), in order to be in an instance of design pattern p, a class has to satisfy at least *cmin* constraints from $\mathcal{R}_p$. After the filtering step, the set of pattern candidates becomes:

$$Cand(S) = S - \{C_j | 1 \leq j \leq n \; s.t. \sum_{i=1, i \neq j}^{n} (1 + |\mathcal{R}_p| - d(C_j,C_i)) < cmin\}$$

2.3 Grouping

After the grouping step we aim to obtain a partition $\mathcal{K} = \{K_1, K_2, ..., K_v\}$ of the set $Cand(S)$ such that each instance of the design pattern p to form a cluster. Based only on the distance semimetric d (Equation (2)), it is possible that two classes would seem to be in an instance of the design pattern (a so called "false positive" decision), even if they are not cohesive enough in order to be grouped together. That is why we need a measure in order to decide how cohesive are two classes. We will consider the dissimilarity degree (from the cohesion point of view) between any two classes from the software system S. Consequently, we will consider the dissimilarity $diss(C_i,C_j)$ between classes C_i and C_j as expressed in Equation (3).

$$diss(C_i, C_j) = \begin{cases} 1 - \frac{|p(C_i) \cap p(C_j)|}{|p(C_i) \cup p(C_j)|} & if \;\; p(C_i) \cap p(C_j) \neq \emptyset \\ \infty & otherwise \end{cases}, \tag{3}$$

where $p(C)$ defines a set of relevant properties of class C and it consists of the application class itself, all attributes and methods defined in the class C, all interfaces implemented by C and the base class of C.

We have chosen the dissimilarity between two classes as expressed in Equation (3) because it emphasizes the idea of cohesion. As illustrated in [2], "*Cohesion refers to the degree to which module components belong together*". Based on the definition of *diss* (Equation (3)), it can be easily proved that *diss* is a semimetric function. In the original paper [8] a theoretical validation of the *semimetric* dissimilarity function *diss* is given. It is proved that *diss* highlights the concept of cohesion, i.e., classes with low distances are cohesive, whereas classes with higher distances are less cohesive.

Consequently, the dissimilarity semimetric *diss* can be used in order to decide how cohesive are two classes. We will use *diss* in the **Grouping** step of our approach in order to decide if two classes are cohesive enough in order to be part of an instance of the design pattern p.

In order to obtain the desired partition $\mathcal{K}$, we introduce a *hierarchical divisive clustering algorithm* (*HDC*).

In our approach the objects to be clustered are the classes from the set $Cand(S)$ and the distance function between the objects is given by the semimetric d (Equation (2)). In the hierarchical clustering process, the dissimilarity semimetric *diss* will be used in order to decide how cohesive are two classes.

The main steps of *HDC* algorithm are:

- A cluster with all the classes from $Cand(S)$ is created.
- The following steps are repeated until the partition of classes remains unchanged (no more clusters can be selected for division):
  - Select the two most distant classes in a cluster from the current partition, i.e., the pair of classes that maximize the distance between them (Equation (2)). If this selection is nondeterministic (there are several pair of classes with the same maximum distance between them), we will choose the pair (C_i, C_j) that has the maximum associated dissimilarity value $diss(C_i, C_j)$. Let us denote by $dmax$ the distance between the most distant classes C_i and C_j from cluster K_r.
  - If $dmax \geq 1 + |\mathcal{R}_p|$ ($|\mathcal{R}_p|$ is the number of constraints imposed by the design pattern p), then cluster K_r will be divided in two subclusters K_{r_1} and K_{r_2}, otherwise the partition remains unchanged. If a division takes place, then cluster K_{r_1} will contain the classes from K_r that are closer (considering the distance d) to C_i than to C_j, and cluster K_{r_2} will contain the remaining classes from K_r. The idea of this step is that a cluster will be divided only if its most distant classes can not be part of an instance of the design pattern p (they invalidate all the constraints that must hold).

2.4 Design pattern instances recovery

The partition $\mathcal{K}$ obtained after the **Grouping** step will be filtered in order to obtain only the clusters that represent instances of the design pattern p. A cluster k from the partition $\mathcal{K}$ is considered an instance of the design pattern p iff the classes from k verify all the constraints from $\mathcal{R}_p$ (the set of constraints imposed by the design pattern p).

3 Experimental Evaluation

In our experiment, we are focusing on identifying instances of the *Proxy* design pattern using the clustering based approach that we have introduced in the previous section.

The *Proxy* design pattern

Proxy is a structural design pattern that provides a surrogate or placeholder for another object to control access to it. Use of proxy objects is prevalent in remote object interaction protocols (*Remote proxy*): a local object needs to communicate with a remote process but we want to hide the details about the remote process location or the communication protocol. The *proxy* object allows to access remote services with the same interface of local processes. In fact, when an *Operation* is required to the proxy object, it delegates the implementation of the required operation to the *RealSubject* object. Being both *Proxy* and *RealSubject* subclasses of *Subject*, this

guarantees that they export the same interface for *Operation*. To be able to call *RealSubject* methods, *Proxy* needs an association to it.

According to our previous considerations, the design pattern *proxy* can be defined as the pair $proxy = (\mathcal{C}_{proxy}, \mathcal{R}_{proxy})$, where: $\mathcal{C}_{proxy} = \{C_1, C_2, C_3\}$; $\mathcal{R}_{proxy} = \{r_1, r_2, r_3\}$, and the constraints are: $r_1(C_1, C_2)$ represents the relation "C_2 extends C_1"; $r_1(C_1, C_3)$ represents the relation "C_3 extends C_1"; $r_1(C_2, C_3)$ represents the relation "C_2 delegates any method inherited from a class C to C_3, where both C_2 and C_3 extend C".

Considering the above, the minimum number of binary constraints *cmin* from $\mathcal{R}_p$ that a class from $\mathcal{C}_p$ can satisfy (as indicated in Equation (1)) is 2.

Example

Let us consider as a case study the simple design illustrated in Figure 1.

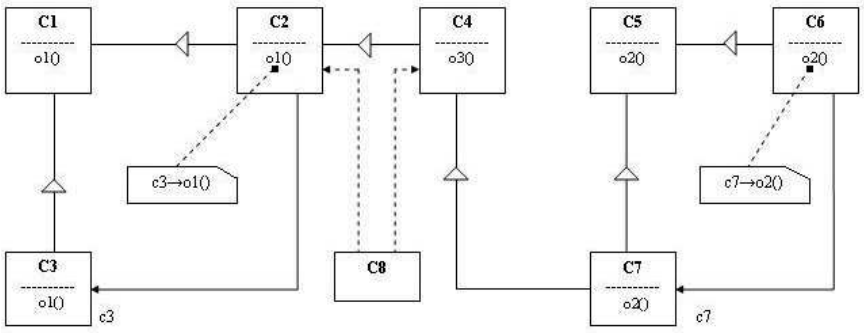


Figure 1: The example design *S*.

For the analyzed design *S*, the set of classes is $Class(S) = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8\}$ and the number of classes is $n = 8$. After applying our approach presented in Section 2, the obtained partition of the software system *S* is $\mathcal{K} = \{K_1, K_2, K_3\}$, where $K_1 = \{C_4\}$, $K_2 = \{C_3, C_1, C_2\}$ and $K_3 = \{C_5, C_6, C_7\}$.

We mention that without using the dissimilarity semimetric *diss*, the class C_7 would have been grouped with the class C_4, instead of being grouped with classes C_5 and C_6 and an instance of the design pattern *Proxy* would have been missed. Now we analyze the obtained partition $\mathcal{K}$ in order to identify instances of *Proxy* design pattern, and the identified instances are correctly reported: K_2 and K_3.

4 Related Work

We will briefly present, in the following, the most significant results obtained in the literature in the field of automatic design patterns identification.

Different approaches, exploiting software metrics, were used in previous works to automatically detect design concepts and function clones [5, 6] in large software systems. An approach for extracting design information directly from C++ header files and for storing them in a repository is proposed in [6]. The patterns are expressed as PROLOG rules and the design information is translated into facts. A single Prolog query is then used to search for all patterns. The disadvantage of this approach is that it handles a small number of design patterns (only the structural design patterns - Adapter, Bridge, Composite, Decorator and Proxy) and the precision obtained in recognition is small (40%).

A multi-stage approach using OO software metrics and structural properties to extract structural design patterns from object oriented artifacts, design, or code, is introduced in [1]. The drawback of this approach is that only few pattern families (the structural design patterns - Adapter, Bridge, Composite, Decorator and Proxy) are approached.

For a precise design pattern recognition, a static analysis is not sufficient. The behavioral aspects of a pattern are an important factor. Dynamic analysis can be used to analyze the runtime behavior of a system. A sole dynamic analysis is not feasible since the amount of data gathered during runtime is too big. A combination of static and dynamic analysis techniques is proposed in [9]. The static analysis identifies candidates for design pattern instances. These candidates form a significantly reduced search space for a subsequent dynamic analysis that confirms or weakens the results from static analysis.

5 Conclusions and future work

We have introduced in this paper a clustering based approach for identifying instances of design patterns in existing software systems.

We can summarize the advantages of the approach proposed in this paper in comparison with existing approaches: the overall worst time complexity ($O(n^3)$) of our approach is reduced in comparison with the worst time complexity of a brute force approach ($O(n^{|\mathcal{C}_p|})$) (as the number of classes contained in a design pattern p is greater or equal to 3); our approach is not dependent on a particular design pattern (it may be used to identify instances of various design patterns, as any design pattern can be described as specified in Section 2); our approach may be used to identify both *structural* and *behavioral* design patterns, as the constraints can express both structural and behavioral aspects of the application classes from the analyzed software system.

Further work can be done in the following directions: improving the **Preprocessing** and **Grouping** steps from our approach; applying the proposed approach on real software systems; extending the proposed approach towards identifying several design patterns; extending the proposed approach towards introducing design patterns in existing software systems.

References

[1] G. Antoniol, R. Fiutem, and L. Cristoforetti, "Using metrics to identify design patterns in object-oriented software", In *Proc. of the Fifth International Symposium on Software Metrics - METRICS'98*, pp. 23–34, 1998.

[2] J. M. Bieman and B.-K. Kang, "Measuring design-level cohesion", *Software Engineering*, Vol. 24(2), pp. 111–124, 1998.

[3] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object Oriented Software*, Addison-Wesley Publishing Company, USA, 1995.

[4] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review", *ACM Computing Surveys*, Vol. 31(3), pp. 264–323, 1999.

[5] K. Kontogiannis, R. de Mori, E. Merlo, M. Galler, and M. Bernstein, "Pattern matching for clone and concept detection", *Automated Software Engineering*, Vol. 3(1/2), pp. 77–108, 1996.

[6] C. Kramer and L. Prechelt, "Design recovery by automated search for structural design patterns in object-oriented software", In *WCRE '96: Proceedings of the 3rd Working Conference on Reverse Engineering (WCRE '96)*, pp. 208–215, Washington, DC, USA, 1996. IEEE Computer Society.

[7] E. Rich and K. Knight, *Artificial Intelligence*, McGraw Hill, New York, 2nd edition, 1991.

[8] G. Serban and I. Czibula, "On evaluating software systems design", *Studia Universitatis "Babes-Bolyai", Informatica*, Vol. LII(1), pp. 55–66, 2007.

[9] L. Wendehals, "Improving design pattern instance recognition by dynamic analysis", In *Proc. of the ICSE 2003 Workshop on Dynamic Analysis (WODA)*, pp. 29–32, 2003.

István Gergely Czibula, Gabriela Şerban
Babeş-Bolyai University
Department of Computer Science
1, M. Kogălniceanu Street, 400084, Cluj-Napoca, Romania
E-mail: {istvanc, gabis}@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 253-258

CELLSIM: An Artificial Life Model Inspired by the Basic Single-Cell Organism

Abbas Pirnia-ye Dezfuli, Bahar Khanahmad Liravi, Mozhdeh Nourizadeh

Abstract: In this paper, we have tried to propose a model for an artificial life system based on the basic biological principles of cells. The two key concepts in this modeling are "cell" and "environment". The cell takes in its vital materials from the environment, and changes its location if necessary. In this simulation, generally, one cell starts its life in the environment, and if the conditions are suitable, the cell reproduces new cells, and so on. Therefore, after a while, a population of cells will be living in the environment, and performing life activities (e.g. feeding, moving and reproduction) until the environment circumstances make them die. There are several parameters related to the cell and environment affecting the simulation such as the amount of food available in the environment and the cell energy consumption due to movement. We report some results of the CellSim, an implementation of this model, and discuss how the environment and the cell parameters affect the simulation.
Keywords: Artificial life, Cell simulation, life activities.

1 Introduction

The phrase "artificial life" was coined by C. Langton: "The ultimate goal of the study of artificial life would be to create 'life' in some other medium, ideally a *virtual* medium where the essence of life has been abstracted from the details of its implementation in any particular hardware. We would like to build models that are so life-like that they cease to be models of life and become *examples* of life themselves." [Langton]. More than two decades later, the debate about artificial life remains very active. Although, to date, some remarkable realizations have been done, such as *Venus* [Rasmussen], *Tierra* [Ray, 1992] or *Cosmos* [Taylor], and pompous declarations such as the famous *How I Created Life in a Virtual Universe* [Ray, 1993], there are still several open problems in artificial life. In a brilliant work [Bedau et al], Bedau and his colleagues listed fourteen open problems in AL, each of which is a grand challenge requiring a major advance on a fundamental issue for its solution. This paper is the result of an investigation concerning three problems of the mentioned Bedau's list.

In this research, we tried to make an artificial life model inspired by the basic biological principles of cell [Majd]. Therefore, we had to study the principal attributes and the key activities of the basic animal cell [Majd] to establish a well-defined model for our artificial cells and their artificial environment.

The paper is organized as follows: section 2 explains "the model" embracing "environment", "cell structure" and "the life process", section 3 presents the "simulation results" and section 4 interprets the results and concludes the paper. In the conclusion, we point out which problems (among Bedau's list) this research deals with.

2 The Model

There are two key concepts in CellSim model: *the environment* and *the cell*. The whole cell life activities take place in the environment which itself is constantly changing through the cell's activities. The cells take in their vital materials from the environment and can move in it to find more suitable locations containing more vital materials [Jawetz]. Performing vital activities and at the same time evolving, the cells may either reach the condition satisfactory for reproduction or come to death due to the environmental circumstances.

2.1 Environment

To begin life simulation, we need an environment in which the process of life can progress. Therefore, we assume a three-dimensional space as the environment partitioned to some clusters. Figure 2 illustrates the *Environment* object components in our simulation.

In addition to the environment clusters, we use an identifier generator that produces a unique number on each call (we will use these numbers as cells' identifiers) and a global time that shows the passage of time in the environment. The clusters contain food, oxygen and waste. The capacity of the clusters is fixed and is equal to the

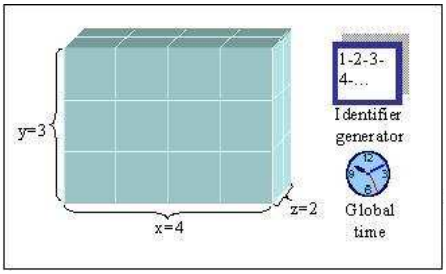


Figure 1: A three-dimensional environment consists of 24 clusters an "identifier generator" and a "global time".

amount of the cluster's oxygen (i.e. the gas which cells use for breathing) plus the food (i.e. the material that cells consume as nutrient) and the waste (i.e. the material that cells produce due to food oxidation).

$$cluster\_O_2 + cluster\_Food + cluster\_Waste = cluster\_Capacity \tag{1}$$

$$Enviroment\_Food = \sum_{i=1}^{x}\sum_{j=1}^{y}\sum_{k=1}^{z}(food\_of\_cluster(i,j,k)) \tag{2}$$

The values of the $cluster\_O_2$, the *cluster_Food* , and the *cluster_Capacity* , in equation (1), are set by the user runtime, and so are x, y, and z in equation (2).

2.2 Cell Structure

The attributes we assume for the cell can be categorized as static and dynamic attributes. The values of the static attributes for each cell are determined when the cell is born, and they never change. Cell identifier, birth place (the number of birth cluster), birth time and maximum size are examples of static attributes. The dynamic attributes are those which are initialized at the birth time, and they change due to the passage of time. Current location, size, food reserve, and the number of ATPs are instances of dynamic attributes.

Any changes in the value of dynamic attributes in the cell are the natural result of operations performed as the cell behavior. This behavior includes *breathing, feeding, movement* and *reproduction*.

***Feeding*:**

The cell consumes some of its ATPs in order to take food from its cluster. Therefore, the number of ATPs in the cell decreases and the amount of food reserve increases, and the amount of food available in the cluster decreases due to the feeding.

***Breathing*:**

In breathing, the cell takes oxygen from its cluster in order to oxidize a part of the cell's food reserve and produce some energy in ATP form. The reaction equation is as follows:

$$\alpha Food + \beta O_2 \rightarrow \gamma Waste + \delta ATP \tag{3}$$

where α, β, γ and δ are coefficients the values of which can be set by the user runtime.

***Movement*:**

The cell departs from the cluster if the amount of oxygen or food is less than specific values that are set by the user runtime. In such a situation, consuming some ATP, the cell moves to an adjacent cluster selected randomly.

***Reproduction*:**

When a cell gets to the user-predefined conditions, the reproduction (i.e. some kind of cell division) takes place. The determining factors of these conditions are the size, the food reserve and the number of ATPs of the cell. One part of the ATP is consumed for reproduction, and the remaining is equally divided up to pass to the two offspring. The cell's food reserve is divided up as well.

2.3 *Life Process*

The process of the cell life is based on its behavior described before (i.e. breathing, feeding, movement, and reproduction). As a cell is born, takes one breath in each time cycle, and the size of the cell increases by a given amount. Thereafter, if the amount of the food reserve in the cell is less than a given part of its food capacity, the feeding begins. Movement and reproduction take place when their conditions are satisfied. The flowchart of the cell life process is illustrated in figure 2.

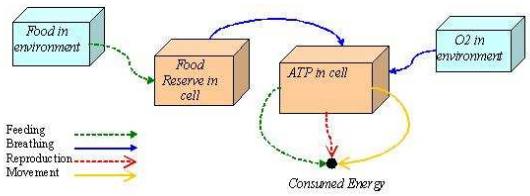


Figure 2: Feeding and breathing affect the environment; movement, reproduction and feeding decrease the ATP in the cell.

This lifecycle terminates because of either reproduction or death. Death, in its turn, may happen through a lack of food or oxygen in the environment.

3 Simulation Results

According to the modeling described in the previous section, we developed an implementation using C# language in .NET framework. In addition to *Cell, Environment* and *Cluster* objects, a *Displayer* object is also defined and added to show the cell and the environment information during the performance. Each cell should perform its vital activities independently. Therefore, each of the cell's life process is carried out by one separate thread, and the thread is aborted when the cell should die.

There are so many parameters that should be set before running such as environment parameters and cell operation parameters. A complete list of all the parameters is shown in table 1. The performance begins with one cell that is located in the central cluster of the environment. It begins its life and evolves until the conditions are suitable for reproduction. Then, it divides up into two new cells. During its life, the cell may move to a new cluster other than its birth location because of a deficiency of food or oxygen. After a while, the processes of life in CellSim i.e. birth, growth and reproduction gradually ends due to a lack of vital materials (i.e. food and oxygen).

Figures 3 and 4 illustrate the results of the program performance using the parameters values brought in table 1. The values of the parameters used in the model are not to be taken as directly comparable to the corresponding values in real cell, because this model is built to show only qualitatively, not quantitatively, adequate behaviour.

4 Interpretation And Conclusion

As it can be seen in figure 3, the amount of food and oxygen decreases as time goes by. In both of them, the rate of decrement increases as population increases and vice versa. The only difference between these two charts is that the food curve is uneven, but that of the oxygen is even. This is because the uptake of the oxygen for breathing is directly and continuously done from the environment, but feeding is done just when the amount of the food reserve is less than a given part of the cell's food capacity.

The population curve depicted in figure 4 implies that the population first increases owing to the good conditions of the environment, and then, when it reaches a certain stage, decreases as a result of vital material deficiency.

Table 1: Parameter values used in presented CellSim run.

	Parameters	Value	Description
Clusters Parameters	cluster capacity	10,000	the total material capacity of each cluster
	initial food volume	4500	the food volume of each cluster at the beginning
	initial O2 volume	5500	the O2 volume of each cluster at the beginning
	environment dimensions	3*3*3	the number of clusters in each dimension i.e. x, y and z shown in figure 1
Cell Attributes	maximum base size	1000	the maximum size that no cell can exceed
	food capacity-size ratio	0.5	the cell food capacity in proportion to its size
	ATP capacity-size ratio	0.5	the cell ATP capacity in proportion to its size
	evolve amount	2	the increase in the cell size in each breathing action at the beginning
	initial base size	500	the size of the cell at the beginning
	initial ATP	100	the amount of the ATP of the cell at the beginning
	initial food reserve	250	the amount of the food reserve of the cell at the beginning
	Comp Q	1	the coefficient of the food capacity to specify the lower limit of the cell food reserve. If the cell food reserve becomes less than the [food capacity * Comp Q], the feeding begins
Cell Operations	ATP consumption for feeding	1*feeding amount	the amount of ATP consumed to feed in proportion to the amount of the food taken in
	ATP consumption for movement	0.05*total size	the amount of ATP consumed for movement in proportion to the total size of the cell i.e. the cell's base size plus the cell food reserve
	ATP consumption for reproduction	0.3*base size	the amount of ATP consumed for reproduction in proportion to the cell size
	α, β, γ and δ	2, 10, 12, and 10	the coefficient of the food, O2, waste, and ATP in equation (3) reproductionin proportion to the cell size

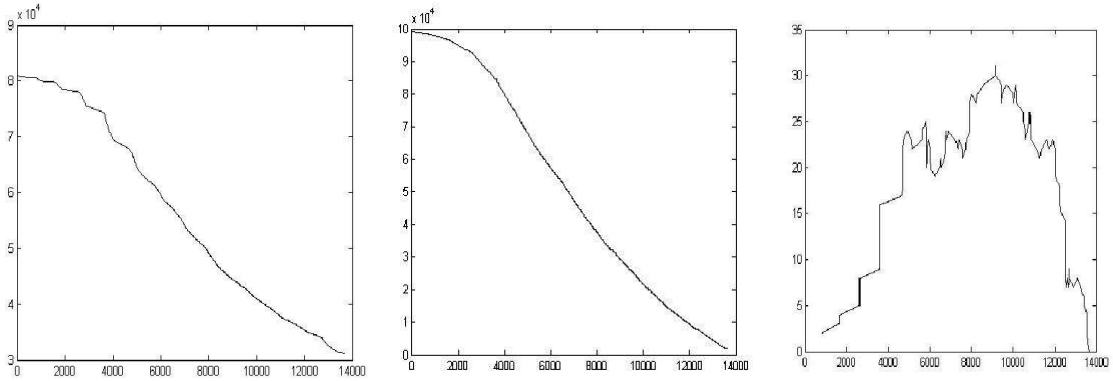


Figure 3: The total food volume (left chart) and oxygen volume (middle chart) in the environment decreases as time passes. The population (right chart) increases first because of cell reproduction, and then decreases because of mortality on account of a lack of vital materials.

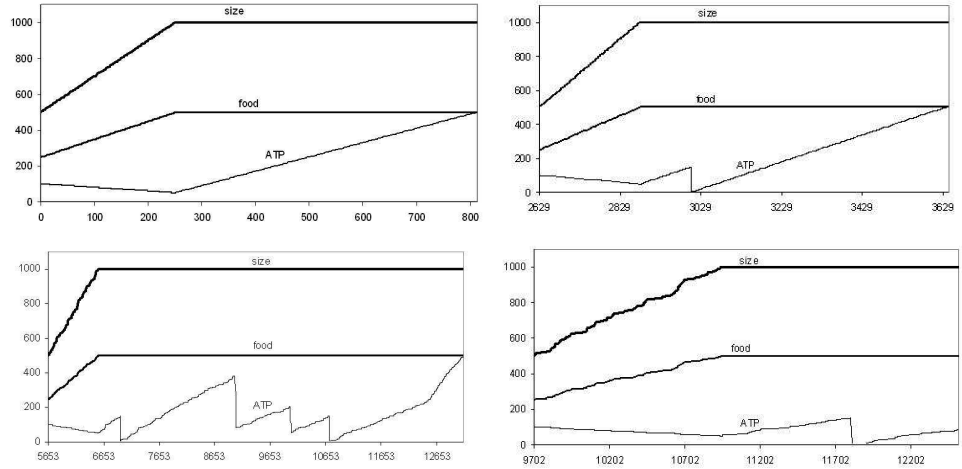


Figure 4: The size, food reserve and ATP of Cells 1, 10, 50, and 100 within their lifetime.

As a comparison between the lifecycles of different cells, it can be seen that the size and the food reserve of cell 1 (fig. 4, top-left) increase easily, because its cluster contains sufficient food and oxygen, and the cell does not have to move to find them. Its ATP decreases as its food reserve increases, because it consumes some energy during the feeding. At the time 813, the reproduction conditions are satisfied and reproduction takes place. The interpretation for cell 10 (fig. 4, top-right) and cell 50 (fig. 4, bottom-left) is the same as the above, but they had to migrate to other clusters to find vital materials. That's why sometimes the ATP curve falls down. Also fig. 4 shows that the number of movements of cell 50 is more than cell 10, because the vital materials available in the environment are rare during its lifetime. In the case of cell 100 (fig. 4, bottom-right), it can be seen that the cell has died at the time 12520 before reaching the reproduction phase because of a lack of vital materials. Finally, the life process in the system ends with the death of the last cell, cell 127, at the time 13674.

As it was mentioned in the introduction, this work is concerned with the three of the fourteen problems of the artificial life put forward by Bedau et al [Bedau]: problems 4, 10 and 11.

4: *Simulate a unicellular organism over its entire lifecycle*: We have described the basic life activities of the cell i.e. birth, growth and reproduction or death.

10: *Develop a theory of information processing, information flow, and information generation for evolving systems*: Figure 2, shows how information are processed during simulation.

11: *Demonstrate the emergence of intelligence and mind in an artificial living system*: In this simulation, the cell doesn't move as long as its current cluster contains oxygen and food, and moves to another cluster selected

randomly as the cluster runs out of food or oxygen. This behavior shows a very low level IQ of the cell. We can enhance the cell IQ by adding some conditional actions to its basic actions, such as going to the cluster that contains the most food and oxygen among its neighbor clusters when it wants to move to another cluster.

References

[1] M. A. Bedau, J. S. McCaskill, N. H. Packard, S. Rasmussen, C. Adami, D. G. Green, T. Ikegami, K. Kaneko, T. S. Ray "Open Problems in Artificial Life", *Massachusetts Institute of Technology, Artificial life 6 (2000):363-367*, 2001.

[2] E. Jawetz, J. L. Melnick, E. A. Adelberg, "Jawetz, Melnick and Adelberg's Mdeical Microbiology", 22 ed, translated into Farsi by J. Nowroozi, pp. 52-63, ISBN 964-460-493-8, Hayyan, 2002.

[3] C. G. Langton "Studying artificial life with cellular automata", *Physica D*, 22, 120-149, 1986.

[4] A. Majd, S. M. A. Shariatzadeh, "Cell and Molecular Biology", 4th edition, pp. 23-29 ISBN 964-7006-35-7, Aeeizh, 2004.

[5] S. Rasmussen, "The Coreworld: Emergence and Evolution of Cooperative Structures in a Computational Chemistry", *Physica D*, 42, 111-134, 1990.

[6] T. S. Ray, "An Approach to the Synthesis of Life", *Artificial Life II*, (pp. 371-408), Redwood City: Addison Wesley, 1992.

[7] T. S. Ray, "How I Created Life in a Virtual Universe", 1993, from http://www.isd.atr.co.jp/ ray/pubs

[8] P. Ritter, "Biochemistry, a foundation", *Brooks/Cole*, p. 301, Pacific Grove CA, 1996.

[9] T. J. Taylor "The Cosmos Artificial Life System", *Information Research Report No. 263*, Department of Artificial Intelligence, University of Edinburg, 1999.

[10] J. Trefil, "1001 Things everyone should know about science", *Doubleday*, p. 93, New York, 1992.

Abbas Pirnia-ye Dezfuli
Azad University-Shiraz Branch
Computer Eng. Department
E-mail: pirnia@iaushiraz.ac.ir

Bahar Khanahmad Liravi
Shahid Beheshti University
Biology Department
E-mail: b.liravi@gmail.com

Mozhdeh Nourizadeh
Azad University-Shiraz Branch
Computer Eng. Department
E-mail: mnzadeh@iaushiraz.ac.ir

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 259-264

Evolutionary Coalition Formation in Full Connected and Scale Free Networks

Laura Dioşan, Dumitru Dumitrescu

Abstract: An optimal clusterization model is studied through an original approach that combines an evolutionary algorithm with the principles of the physical spin systems. The method is used in order to investigate the process of coalition formation that appears in complex systems, actually in full connected and scale free networks. The numerical experiments show that the proposed evolutionary model is able to detect the optimal coalition formation in small and large systems by a reasonable cost of complexity (considered in terms of time and physical computational resources).
Keywords: Coalition formation, Complex systems, Evolutionary optimisation.

1 Introduction

Almost all interesting processes in nature are highly cross-linked. In many systems, however, we can identify the elements that interact to form compound structures or functions. This process of emergence of the need for new, complementary, modes of description is known as hierarchical self-organization, and systems that observe this characteristic are defined as complex [1]. Examples of these systems are the brain, the immune system, the biological cells, the metabolic networks, the ant colonies, the Internet and World Wide Web, the economic markets, or the human social networks.

More formally, a complex system is any system featuring a large number of interacting components (agents, processes, etc.) whose aggregate activity is nonlinear (not derivable from the summations of the activity of individual components) and typically exhibits hierarchical self-organization under selective pressures [2].

An interesting problem that appears in such complex systems is the process of coalition formation. The optimization of several models proposed in order to study the coalition formation in complex system (politics, economics or sociological systems) are based on simulated annealing [3] or on extremal optimization method [4]. Both optimization techniques are rather time-consuming and the necessary computational time increases strongly with the system size. The evolutionary techniques overtake these weaknesses. It will be shown in this paper that the evolutionary methods can simulate very well the true dynamic of such complex systems and they allow analyse the phase transition from the viewpoint of the much-discussed social percolation [5], where the emergence of a giant cluster is observed in many social phenomena.

Physical concepts might prove useful in describing collective social phenomena. Indeed, the models inspired by statistical physics are now appearing in scientific literature [6]. The process of aggregation among a set of actors seems to be a good candidate for a statistical physics like model [7]. These actors might be countries, which ally into international coalitions, companies that adopt common standards, parties that make alliances, individuals that form different interest groups, and so on. Given a set of actors, there always exists an associated distribution of bilateral propensities towards either cooperation or conflict. The question then arises as to: *How to satisfy such opposing constraints simultaneously?*. In other words, *what kind of alliances, if any, will optimize all actor bilateral trends to respectively conflict or cooperation?*

It turns out that a similar problem does exist in spin glasses (as Ising or Potts models [8]). For these systems, the magnetic exchanges are distributed randomly between the ferro and anti-ferromagnetic couplings. Indeed such an analogy has been used in the past in a few models [7].

The aim of this paper is to develop the hybrid model proposed in [9] in order to investigate the process of coalition formation that can appear in two network types: full connected and scale free networks. Coalition setting among a set of actors (countries, firms, individuals) is studied using concepts from the theory of spin glasses and from the theory of evolution. Those of evolutionary computation combine the principles of the Potts model. Unlike other solutions proposed until now in order to study the dynamic of such complex system (simulated annealing or Monte Carlo methods), the proposed model is able to deal with large systems and helps to investigate different phenomena that appear in such systems. The numerical results indicate that the evolutionary approach is able to identify the characteristics of the coalition formation process.

The rest of the paper in organised as follows. The Potts model is briefly described in Section 2. Section 3 proposes the new evolutionary approach in order to investigate the process of coalition formation. Several numerical experiments are presented and discussed in Section 4. Finally, the last section concludes the paper.

2 The Potts model

Understanding human thinking and learning has always been a great challenge for all the scientists and not only. The challenge has taken another dimension as scientists are trying to simulate the learning and thinking processes by using the computers and other devices. The nature inspired and Physics models have been of great help.

Even though very simple, the Ising model and its generalization, the Potts model, have been applied successfully in several computational problems. In its original form, the Ising model describes the evolution of a grid of up and down spins over time. Each spin can change its orientation in time, according to the external temperature and the values of its orthogonal neighbours [10]. The Potts system involves similar dynamics for the spins, but each spin can have more than two (up and down) orientations.

A simple version of a spin glass [11] consists of a d-dimensional hyper-cubic lattice with a spin variable $\sigma_i \in \{-1,1\}$ placed on each site i, $1 \leq i \leq n$. A spin is connected to each of its neighbours j via a bond variable $J_{i,j}$ drawn from some distribution $P(J)$ with zero mean and unit variance [7].

The infinite-range p-state Potts glass is usually defined by the Hamiltonian: $H(\sigma) = -p\sum_i \sum_j J_{ij}\delta_{\sigma_i\sigma_j}$, where the $\sigma(i)$ Potts states can take the $0,1,2,\ldots,p-1$ values. The sum is extended over all $N(N-1)/2$ pairs and $\delta_{mn} = 1$ if $m = n$ and $\delta_{mn} = 0$ otherwise. The J_{ij} bonds are randomly distributed quenched variables with J_0/N mean, and the variance is presumed to scale as N^{-1}. The system is non-trivially frustrated and computing the thermodynamic parameters is a complex task. The above model has been extensively studied by many authors through different methods [8]. The main idea of these models is to find the "ground states", *i.e.*, the lowest energy configuration S_{min} of the Hamiltonian.

Neda et al. have considered a model resembling the infinite-range Potts glass [7], which can be useful for considering the optimal clusterization problem or for understanding the coalition formation phenomena in sociological systems. A difference to the Potts glass is that now the variance of the J_{ij} bonds scales as N^{-2}. The authors [7] have considered an unrestricted number of Potts states ($p = N$), and limit the study on the ground state ($T = 0$).

Therefore, this non-trivial optimization problem can be mathematically formulated resembling a zero-temperature Potts glass type model. To prove this, a cost-function K (a kind of energy of the system) has been defined. This function has been increased by $S_iS_j|Z_{ij}|$ whenever two conflicting actors (i and j) are in the same coalition or two actors which have a tendency towards collaboration are in different coalition. The cost-function is zero, when no propensity is in conflict with the formed coalitions. The number of possible coalitions is unrestricted (the maximal possible number is N), and the coalition in which actor i is denoted by $\sigma(i)$ [7]. The cost function then writes as

$$K = -\sum_{i<j} \delta_{\sigma(i)\sigma(j)} Z_{ij}S_iS_j + \frac{1}{2}\sum_{i<j}(Z_{ij}S_iS_j + |Z_{ij}S_iS_j|) \tag{1}$$

The order parameter considered by Neda et al. in [7] has been the relative size r of the largest cluster:

$$r = \left\langle \overline{max_i \left\{ \frac{C_x(i)}{N} \right\}} \right\rangle_x, \tag{2}$$

where $C_x(i)$ stands for the number of elements in state i for an x realization of the system [7].

3 The evolutionary-based coalition model (ECM)

3.1 Representation

An evolutionary model is developed in order to study the coalition formation process in complex systems, which are considered as Potts glass systems. The ECM evolves the realizations $\sigma(i)$ of a network/system configuration who's energy has to reach a minimal value. Actually, each ECM individual is a fixed-length string of genes. The length of a chromosome is equal to the number of nodes from the network. Thus, an ECM chromosome represents a possible clusterization of the network nodes in order to form an optimal coalition. Because the maximal number of coalitions (or clusters) is equal to the number of network's nodes, each gene is associated to the index of such a cluster. Therefore, each gene is an integer number from $\{1,2,\ldots,N\}$ set (where N represents the number of nodes from the network). Or, in terms of the Potts model, each gene g_i from an ECM chromosome is associated to a Potts state $\sigma(i)$.

As regards the chromosome initialization, each gene of a chromosome is initialized with a random value from the $\{1, 2, \ldots, N\}$ set; in this case two or more nodes could take part to the same coalition from the start of the search process.

3.2 Fitness assignment

The array of integers encoded into an ECM chromosome represents the structure of a coalition. In order to compute the quality of a coalition, the cost function proposed by Neda et al. [7] is used. The simple case when $S_i = S_j = 1$ and $Z_{ij} = +1$ with a probability q and -1 with a probability $1-q$ is considered:

$$f = -\sum_{i,j=1}^{N} Z_{ij} \times \delta_{g_i,g_j}, \text{ where: } Z_{ij} = \pm 1, \text{ and } \delta_{g_i,g_j} = \begin{cases} 1, & \text{if } g_i = g_j \\ 0, & \text{if } g_i \neq g_j \end{cases} \tag{3}$$

A lower value of this function indicates a better quality of the chromosome. Therefore, the ECM has to solve a minimization problem.

3.3 Search operators

The search operators mainly used within the ECM are the crossover and mutation. Note that the action of the genetic operators does not change the structure of the network (the interactions between the actors). The crossover and the mutation change the coalitions only, which are formed in the system.

Crossover. By crossover, two selected parents are recombined. For instance, within the cutting-point recombination, two possible coalition configurations (one from each parent) exchange the elements placed between the cutting-points. A cutting point is considered within the following two parent chromosomes.

Mutation. By mutation, some information inside a chromosome could be changed. In other words, some of the network actors could change their group affiliation. Therefore, by mutation, a gene change its value into another one (off course, from the same discrete domain $\{1,2,\ldots,N\}$).

3.4 The evolutionary mechanism

The steady-state evolutionary model is used as an underlying mechanism for the ECM implementation. The algorithm starts by creating a random population of individuals. The following steps are repeated until a given number of generations is reached: two parents are selected by using a binary tournament selection procedure. The parents are recombined in order to obtain two offspring by performing a one cutting-point crossover. The offspring are considered for mutation. The best offspring O replaces the worst individual W in the current population if O is better than W.

4 Numerical experiments

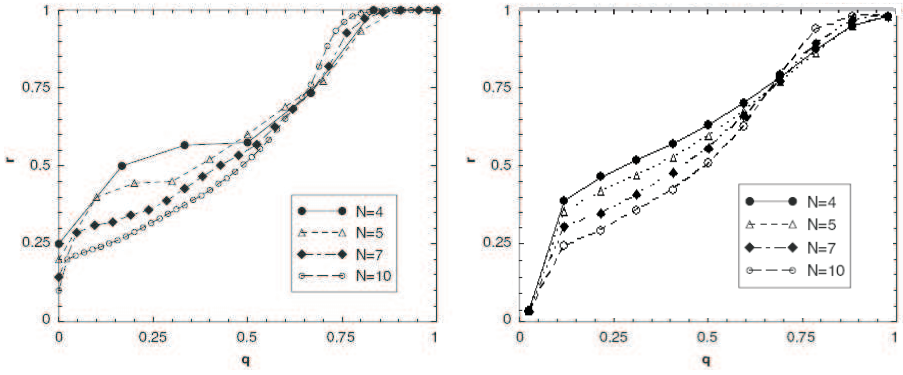
In this experiment the dynamic of the coalition formation through the order parameter r (like in [7]) is studied in full connected networks - there is a positive connection (a sympathy) or a negative connection (an antipathy) between every 2 nodes of the network. In addition to this study, the results are compared to those computed by using Monte Carlo methods for small full connected systems (up to $N \leq 10$). Several numerical experiments are also presented for large full connected systems (e.g. $N = 50$ or $N = 100$).

4.1 ECM in full connected networks

First, the ECM is used in order to obtain the optimal coalition formation for small systems in which an exact enumeration is possible. The exact enumeration means that one can computationally map the whole phase-space (all $\sigma(i)$ realizations) for a generated Z_{ij} configuration and determine the minimum energy state. The order parameter considered is the relative size r of the largest cluster.

Moreover, for $N \leq 7$ it was also possible to map all the Z_{ij} configurations as well. The results from [7] up to $N \leq 7$ are thus exact. In the $7 < N \leq 10$ interval, although the minimum energy states are exactly found, due to greatly increased computational time and memory needed, it was possible to generate only a reasonable ensemble

Figure 1: Results of the dependence of the order parameter as a function of q for different sizes of the network (N). For comparison purposes on (a) the exact enumeration results are shown [7] and on (b) the ECM optimisation results.



averaged for Z_{ij} (5000 configurations) [7]. The results obtained by the evolutionary approach proposed in this paper up to $N \leq 10$ are averaged over the same reasonable ensemble of 5000 network configurations.

The comparison exact enumeration is performed of two purposes. First, the trends of the $r(q)$ curves as a function of increasing system size is checked. Secondly, these results offer a good "standard" for the proposed evolutionary optimisation method, used for larger system sizes (in the next experiment). As the results in Figure 1 show, the $r(q)$ curves have a similar trend as those suggested by Neda et al. in [7], i.e., as the system size increases, the slopes for $r(q)$ are increased around a non-trivial q value.

The ECM results are in perfect agreement with those from exact enumerations [7], giving confidence in to use the evolutionary optimisation model. In addition, the complexity of the proposed approach is smaller than the complexity of the traditional methods, which have been applied in order to investigate the coalition formation process [7]. Therefore, the time that is needed in order to identify the optimal clusterization of the "actors" in such system by the evolutionary methods is reduced. Another and maybe the most important advantage of the proposed approach is given by its ability of handle non-linear, high dimensional problems without requiring differentiability or explicit knowledge of the problem structure. This characteristic is very important and it favours a new direction in studying the phenomena that appear in very large complex systems.

In [7] the authors have considered two Monte Carlo optimisation methods: the classical simulated annealing [12] and the recently proposed extremal optimization method [4, 13]. Both approaches are rather time-consuming and the necessary computational time increases sharply with the system size. The computational resources allowed the authors to study only the systems by sizes up to $N \leq 60$.

The evolutionary approach studied in this paper is time-consuming also, but the computational resources allow investigating the coalition formation in systems by larger size. Therefore, the evolution of the order parameter $r(q)$ is analysed in four large systems with $N = 25$, $N = 50$, $N = 75$ and $N = 100$, respectively, nodes with a statistic of 100 realisations.

Even if for the large systems, large populations are evolved during more generations than those used for the small systems, the computational time that is needed in order to obtain good solutions is reasonable.

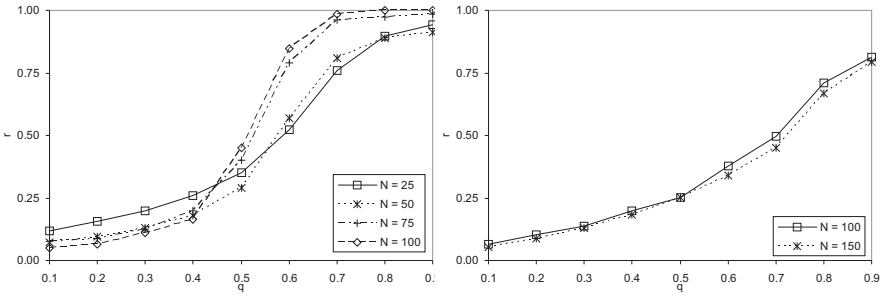
4.2 ECM in scale free networks

In this experiment the evolutionary model is used in order to study the coalition formation in scale free networks with different number of nodes [14]. The BA model [15] is used in order to generate each scale free network. The evolution of the same order parameter r is again investigated. A population of 5000 individuals is evolved during 5000 generations. The results obtained for two network dimensions ($N = 100$ and $N = 150$) and for different values of the probability for a "sympathy" connection are presented in Figure 2.

4.3 Discussions

Several aspects can be remarked from the numerical results presented in Figures 1 and 2:

Figure 2: Results of the dependence of the order parameter as a function of q for: a) large full connected systems and b) scale free networks. The optimal values of the order parameter r are evolved by the ECM approach. The results are averaged over 100 configurations for each value of q.



- when in the system there are more relations of conflict than the collaboration ones ($q \to 0$), in the full connected networks usually the nodes tend to form as much clusters as nodes are (the nodes tend to form the one's own cluster), while in the scale free networks, the number of clusters is large, but less to the number of nodes.

- when in the system there are more relations of collaboration than the conflict ones ($q \to 1$), usually the nodes from a full connected network tend to form a single cluster in order to satisfy the conflicting interactions, while the nodes from the scale free networks tend to form a small number of clusters, but not only one (this fact can be translated by the presence of the hubs that "control" some regions from a scale free network).

- while the equilibrium state of the system represents a break-point for the full connected networks, in the scale free cases the largest changes appear for a value of q closed to 0.7 (in other words, there are needed of more collaborations connections in order to change the structure of the optimal coalition).

5 Conclusions

An evolutionary framework has been proposed in this paper in order to study the process of coalition formation that appears in complex systems. Two types of networks have been investigated: full connected and scale free networks. Two types of full connected networks have been considered: small networks (up to 10 nodes) and large networks (from 10 up to 100 nodes), which are closer to the real systems than the smaller ones. The numerical results obtained in all cases indicate the relationship between the number of coalitions and the structure of the network.

The proposed model presents several advantages: the time that is needed in order to identify the optimal clusterization of the "actors" in such systems by the evolutionary methods is reduced; the ability of handle non-linear, high dimensional problems without requiring differentiability or explicit knowledge of the problem structure. This characteristic is very important and it favours a new direction in studying the phenomena that appear in very large complex systems.

Future works will be focused on the study of the coalition formation in full connected networks in which the relations (of collaboration or conflict) between the elements are weighted (instead to have only $+1$ or -1 links between 2 nodes, some fuzzy relations will be defined on $[-1, 1]$ range). The optimal clusterization will be also investigated in networks that are more sophisticated: random networks, small world networks, and scale-free networks.

Acknowledgements

The present study was supported by the National Research Grant "Developing and optimisation of hybrid methods based on evolutionary techniques. Applications for NP-complete optimisation problems" – CNCSIS, Romania. We also thank to prof. Zoltan Neda for his interesting discussion.

References

[1] Pattee, H.: Complementarity principle in biological and social structures. J. Social and Biological Structures **1** (1978) 1–10

[2] Mitchell, M.: Complex systems: Network thinking. Artif. Intell **170**(18) (2006) 1194–1212

[3] Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science, Number 4598, 13 May 1983 **220, 4598** (1983) 671–680

[4] Boettcher, S., Percus, A.: Nature's way of optimizing. Artificial Intelligence **119**(1–2) (2000) 275–286

[5] Solomon, S., Weisbucha, G., de Arcangelisc, L., Janc, N., Stauffer, D.: Complex systems: Network thinking. Physica A **277**(1-2) (2000) 239–247

[6] S. Moss de Oliveira, P.d.O., Stauffer, D.: Evolution, Money, War, and Computers-Non-Traditional Applications of Computational Statistical Physics. Teubner (1999)

[7] Néda, Z., Florian, R., Ravasz, M., Libál, A., Györgyi, G.: Phase transition in an optimal clusterization model. Physica A **362** (2006) 357–368

[8] Erzan, A., et al.: The infinite-ranged potts spin glass model. J. Phys. C: Solid State Phys. (16) (1983) 555–560

[9] Dioşan, L., Dumitrescu, D.: Evolutionary coalition formation in complex network. Studia Universitas Babes–Bolyai, Seria Informatica **LII**(2) (2007) 115–129

[10] Bak, P.: How nature works: The science of self-organized criticality. Springer-Verlag (1996)

[11] Mézard, M., Parisi, G., Virasoro, M.A.: Spin Glass Theory and Beyond. World Scientific (1987)

[12] Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**(4598) (1983) 671–680

[13] Boettcher, S., Percus, A.G.: Extremal optimization for graph partitioning. CoRR **cond-mat/0104214** (2001)

[14] Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439) (1999) 509–512

[15] Barabási, A.L.: Linked: The New Science of Networks. Perseus (2002)

Laura Dioşan[1,2], Dumitru Dumitrescu[1]
[1]Babeş-Bolyai University
Department of Computer Science, Faculty of Mathematics and Computer Science
M. Kogalniceanu nr. 1, RO-400084 Cluj-Napoca, Romania
[2]Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes, EA 4108
Institut National des Sciences Appliquées, Rouen, France
E-mail: {lauras, ddumitr}@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 265-270

Examination of Fault Tolerance in MMPI

Daniel C. Doolan, Sabin Tabirca

Abstract: The Mobile Message Passing Interface (MMPI) provides the developer with a set of functions similar to that found in the Message Passing Interface used on high end parallel machines and clusters. Unlike these specially built machines that feature high speed interconnects the MMPI system is designed on top of Bluetooth technology allowing for parallel applications to be developed within the realm of mobile computing. MMPI is a Java based library built upon Java Micro Edition (JME) and JSR-82 Bluetooth. Fault Tolerance is especially important in MPI as the probability for failure within the system increases as the number of nodes increase. This paper looks at the aspects of fault tolerance from the mobile perspective through the use of the MMPI library. Mobile fault tolerance opens up a whole other dimension of possible faults that need to be handled when dealing with wireless communications technology such as Bluetooth. This paper discusses some of the aspects of Mobile fault tolerance within the MMPI world.
Keywords: MMPI, Bluetooth, Fault Tolerance, Mobile Parallel Computing

1 Introduction

True fault tolerance is something that we can only aspire to. No matter how well a system may be designed their is always the possibility of a complete failure of all nodes within the parallel world. Fault tolerance within the world of mobile parallel computing is far more complex. Murphy's law succinctly defines that if something can go wrong it will go wrong.

It is only within the past few years that the concept of mobile grid / parallel computing has come into it's own. This is due to the lack of system resources inherent with mobile devices. The mobiles of today however sport microprocessors of 220Mhz and better [4] [12] with tens of megabytes of system memory available for application usage. One can expect the mobiles of tomorrow to have gigahertz level processors as a 1Ghz system was announced by ARM [2] [3] [13] [14] in October 2005. Mobiles are advancing at such a rate that in October 2007 Sun Microsystems announced [16] that they would no longer be supporting JME development as mobile devices are on the verge of being capable of running full blown JVM's. It is expected that within ten years most devices will be running such VM's although lower end devices may still be running with the JME VM.

The sheer number of mobile devices currently active is in excess of 3.3 Billion (50% of the worlds population). For the last few years sales of mobile devices have remained constant at approximately one Billion units [10] per year. The combined computing power of all these mobile devices represents a significant computing resource.

1.1 Bluetooth

Bluetooth operates within the range of 2.40Ghz to 2.48Ghz of the Industrial, Scientific and Medical (ISM) RF band. This band is divided into 79 segments displaced by 1Mhz graduations with a maximum frequency hop rate of 1,600 hops/s. The Bluetooth 1.2 specification allows for transmission speeds of up to 721 kbits/s, however the Bluetooth 2.0 specification allows for an Enhanced Data Rate (EDR) of 3.0 Mbits/s giving an effective rate of up to 2.1 Mbits/s. The majority of Bluetooth enabled phones allow for the lower data rate, but some of the more recent ones (Nokia N810 (announced October 2007)) provide Bluetooth 2.0 + EDR, as well as WLAN support 802.11b/g. With the adoption of version 2.1 + EDR on the 1st August 2007 by the Bluetooth Special Interest Group (SIG) and Ultra Wide Band (UWB) Technologies, Bluetooth is clearly here to stay for the near to medium future. Three main classes of Bluetooth network exist: Point to Point, Piconet and Scatternet. The majority of Bluetooth networks that are presently formed are of the Point to Point variety. Hands free wireless headsets being one of the main application sectors. The Piconet allows for the formation of a network consisting of several devices, with an upper limit of eight (a server connected with seven clients). In reality some devices that support Bluetooth do not fully comply and therefore the upper limit can be far more limited. Networks formed using the Piconet configuration have a Star network topology, therefore all inter client communication must be routed through the server device.

1.2 Mobile Message Passing Interface

MMPI [7] allows for the creation of a parallel world through the creation of a fully interconnect mesh network. The mesh is established in a phased manner, firstly a selection of devices are started as client nodes and one other started as a master / root node. The master carries out device and service discovery to find all of the other devices that have advertised themselves as "mmpiNode"s. At this stage the network topology that is formed is that of the standard star network typical of Bluetooth Piconets. According to the Bluetooth specification the inquiry phase must last for 10.24 seconds [5]. However in reality this figure is usually several seconds more. The second phase of the formation algorithm results in the creation of communication channels between all of the "Client" devices. This requires the Client designated devices to create Server connections, on instantiation of same a message is relayed to the master which in turn relays it to the appropriate Client node. On receipt of the relayed message the Client can then establish a connection with the Server object of initiating Client node.

The MMPI library allows for a selection of point-to-point and global communication methods to be called. A developer using the library has no need to develop any Bluetooth code. One must simply instantiate an MMPI object and then call the required communication routines upon same. The ability to develop a single application rather than separate Client and Server applications helps to reduce development time and cost as well as simplifying the system architecture. The library is suitable not only for mobile parallel processing applications but also for multi-player gaming, mLearning and mobile parallel graphics.

2 The Costs for Fault Tolerance in MPI

The Message Passing Interface in itself is simply a standard that specifies how a correctly written parallel program may be achieved. Many say that MPI is not fault tolerant, but this is neither true or false. Most believe that when a process dies then all MPI nodes that are part of the world should so too die, this however is not the case. The default operation for when a process becomes unavailable is indeed to kill all remaining processes in the world, this is achieved through the built in MPI_ERRORS_ARE_FATAL error handler. Therefore if the developer carries out no error handling in respect to this type of error, then indeed all other nodes in the world will die before the process should normally terminate by calling the MPI Finalize function. This default behaviour for all nodes dying on a node having an error was decided by the MPI Forum to be the most useful default behaviour. Fault tolerance is not an actual property of MPI itself. In general it is assumed that that the parallel program will execute on reliable hardware. How hardware faults are handled are implementation specific.

Gropp and Lusk [9] investigated fault tolerance in MPI by dealing with only one probability for a failure of the system. It is assumed that at most one failure may occur between checkpoints with the probability α. Therefore the total run time may be defined by

$$E_T = \frac{T}{t_0}\left(k_0 + t_0 + \alpha\left(k_1 t_0 + \frac{1}{2}t_0^2\right)\right),$$

where k_0 is the time to create a checkpoint and k_1 is the time to read / restore a checkpoint. Accordingly, the optimal time between checkpoints is given by $t_0 = \sqrt{\frac{2k_0}{\alpha}}$ therefore the expected computation time is $T(1+\alpha k_1 + \sqrt{2\alpha k_0})$.

In general a fault tolerant program and underlying infrastructure should be capable of surviving failures such as system crashes and network failures. At the highest level the MPI program should be capable of automatically recovering from a set of faults without any change to the apparent behaviour of the program. The next level is that notifications of failures should be posted to the MPI program so appropriate action may be undertaken. At the third level, certain operations may become invalid, for example failure of a node may rule out the possibility of collective communication routines, but standard point to point communication may still continue between unaffected nodes. The fourth level makes use of checkpointing thereby allowing a program to save its state to persistent storage, abort and restarted from the checkpoint. The final level of survival may use a combination of the previous approaches. Most approaches to fault tolerance have a set of three distinct requirements: detection of failures, the maintaince of state information to continue the computation and the ability to restart. Any implementation conforming to the MPI standard is responsible for detecting and handling network faults, this may include message retransmission or the informing of the application through the use of an error code. Essentially the contents of a message transmitted from one node should be identical to the message received on another node.

Several fault tolerant MPI implementations are currently in existence. MPICH-V [6] is considered to be one of the most complete featuring checkpointing and message logs to allow aborted processes to be replaced. This implementation comes at a cost of approximately doubling the communication times, for the provision of full

recovery. The LAM based MPI-FT [11] also uses a similar approach to fault tolerance while FT-MPI [8] modifies some of the standard MPI semantics.

The first attempt at the development of fault tolerant MPI applications made use of checkpointing and roll back. Co-Check MPI [15] was the first MPI implementation built that used the Condor [17] library for checkpointing. All process would synchronously checkpoint, this proved to be a drawback with the system as in large systems the procedure could become expensive from a time concern. The result of this work was the creation of a new version of MPI called tuMPI as the modification of the original MPICH implementation was considered too complex. Another similar implementation is Starfish MPI [1] but uses its own system to achieve checkpointing. The use of atomic group communication calls removed the need as in the tuMPI implementation to flush the message queues to avoid messages being lost.

3 Quantizing the Costs for Mobile Fault Tolerance

Checkpoints may be created at regular intervals to save program state. If T is the total execution time without checkpoints, and t_0 is the time between checkpoints then the number of checkpoints is given by $\frac{T}{t_0}$. Under standard MPI characteristics a node may fail due to errors on the device itself, or due to errors related to inter-device communications. In the mobile world these failures have been classified into three distinct categories.

1. Normal failure with the probability α_0.
2. A device fails because it exceeded the Bluetooth range. This occurs with the probability α_1.
3. A device terminates the application with the probability α_2.

The accurate detection of errors is only one half of the fault tolerance equation, the other is to ensure that applications can carry on from a previously valid system state. This may be achieved through the use of checkpointing. In the case of errors caused by devices moving in and out of the Bluetooth range (10 meters, for a typical phone), one may attempt to restore the original connections (DataInput / DataOutput Streams) as the physical address of the devices in question remain the same. In other cases it may be necessary to re-initialise the world. This would require for the carrying out of device and service discovery once again, and the reformation of the network based on the presently active nodes detected by the discovery process. This can be an expensive operation as the combined discovery process can last on the order of eighteen to twenty seconds.

$k_0 =$ the time to create a checkpoint
$k_1 =$ the time to read / restore a checkpoint
$k_2 =$ the time to restore the communication channels
$k_3 =$ the time to initialise the world

3.1 Evaluating the Costs

There are three distinct cases to be evaluated as mentioned previously with regard to possible failures. The overall cost of fault tolerance is a combination of these three together.

Case 1. This is when a normal failure occurs. The costs involved are to create a checkpoint, read the previous checkpoint and restore the state hence the equation is exactly as in Gropp & Lusk [9]. For one checkpoint we have

$$E_1 = (1-\alpha_0 t_0)(k_0+t_0) + \alpha_0 t_0 \left(k_0 + t_0 + k_1 + \frac{1}{2}t_0\right) =$$
$$= k_0 + t_0 + \alpha_0 \left(k_1 t_0 + \frac{1}{2}t_0^2\right) = k_0 + t_0(1+\alpha_0 k_1) + \frac{1}{2}\alpha_0 t_0^2$$

which gives the following cost over $\frac{T}{t_0}$ checkpoints

$$E_1^t(t_0) = \frac{T}{t_0}\left[k_0 + t_0(1+\alpha_0 k_1) + \frac{1}{2}\alpha_0 T_0^2\right]. \tag{1}$$

Case 2. When a device is outside of the Bluetooth network range. The device should return to within the network range and consequently reestablish the I/O connections with the MMPI world. The device will then read the previous checkpoint and restore the system state. Therefore, the total cost for one checkpoint is

$$E_2 = (1-\alpha_1 t_0)(k_0+t_0)+\alpha_1 t_0\left(k_0+t_0+k_2+k_1+\frac{1}{2}t_0\right)=$$
$$=k_0+t_0+\alpha_1 t_0(k_2+k_1)+\frac{1}{2}\alpha_1 t_0^2 = k_0+t_0\left[1+\alpha_1(k_2+k_1)\right]+\frac{1}{2}\alpha_1 t_0^2$$

with the total cost given by

$$E_2^t(t_0)=\frac{T}{t_0}\left[k_0+t_0\left[1+\alpha_1(k_2+k_1)\right]+\frac{1}{2}\alpha_1 t_0^2\right] \tag{2}$$

Case 3. When the device terminates the application for some reasons. In this case all the devices should start the application from scratch which gives the following total cost.

$$E_3 = (1-\alpha_2 t_0)(k_0+t_0)+\alpha_2 t_0(k_0+t_0+T) = k_0+t_0+\alpha_2 t_0 T. \Rightarrow$$

$$E_3^t(t_0)=\frac{T}{t_0}\left[k_0+t_0(1+\alpha_2 T)\right]. \tag{3}$$

Total cost over these three cases is

$$E^t(t_0)=E_1^t(t_0)+E_2^t(t_0)+E_3^t(t_0)=$$
$$=\frac{T}{t_0}\left[3k_0+t_0\left[3+\alpha_0 k_1+\alpha_1(k_2+k_1)+\alpha_3 T\right]+\frac{1}{2}(\alpha_0+\alpha_1)t_0^2\right]=$$
$$=\frac{3Tk_0}{t_0}+\frac{T}{3}(\alpha_0+\alpha_1)t_0+\left[3+\alpha_0 k_1+\alpha_1(k_2+k_1)+\alpha_2 T\right]T.$$

The optimal time t_0 between checkpoints can be calculated as follows

$$\frac{dEt}{dt_0}=-\frac{3Tk_0}{t_0^2}+\frac{T}{2}(\alpha_0+\alpha_1)=0\Rightarrow\frac{3Tk_0}{t_0^2}=\frac{T}{2}(\alpha_0+\alpha_1)\Rightarrow t_0^2=\frac{6k_0}{\alpha_1+\alpha_2}\Rightarrow t_0=\sqrt{\frac{6k_0}{\alpha_1+\alpha_2}}$$

which gives the following optimal run time

$$E^t(t_0)=\frac{3Tk_0}{\sqrt{\frac{6k_0}{\alpha_1+\alpha_2}}}+\frac{T}{2}(\alpha_1+\alpha_2)\sqrt{\frac{6k_0}{\alpha_1+\alpha_2}}+T\left[3+\alpha_0 k_1+\alpha_1(k_2+k_1)+\alpha_2 T\right]=$$
$$=\sqrt{\frac{3}{2}}\sqrt{k_0(\alpha_1+\alpha_2)}T+\sqrt{\frac{3}{2}}\sqrt{k_0(\alpha_1+\alpha_2)}T+T\left[3+\alpha_0 k_1+\alpha_1(k_2+k_1)+\alpha_2 T\right]=$$
$$=\left[\sqrt{6k_0(\alpha_1+\alpha_2)}+3+\alpha_0 k_1+\alpha_1(k_1+k_2)+\alpha_2 T\right]T.$$

$E^t(t_0)$ represents the minimal time that may be achieved with mobile fault tolerance. Consequently as α_0, $\alpha_2 \simeq 0$ the highest probability for error is that of a device moving outside of the range α_1 therefore the optimal time in this case may be given by $E^t(t_0)=\left[\sqrt{6k_0\alpha_1}+3+\alpha_1(k_1+k_2)\right]T$.

3.2 Saving System State

The saving of checkpoints can be carried out in several different manners depending on the type of application and amount of data required for checkpointing. Firstly, the checkpoint data can be saved to the file system using JSR-75. This however requires the Midlet to be signed, as unsigned Midlets require significant user intervention to give the application permission. This means of checkpointing is necessary for large and complex applications where a large amount of system state must be saved. The alternative to this is to store the state data using the Record Management System RMS. No user intervention is required for this, but the RMS is only suitable for applications where a small amount of state information is necessary. The final checkpoint process allows for the saving of system state to the programs internal memory, this provides the fastest checkpointing facility. This checkpoint is volatile but can be used in conjunction with the persistent forms to speedup the recovery time.

4 Summary and Conclusions

Parallel computing within the mobile domain opens the door the to a plethora of new possibilities for inter-device communications failure. True fault tolerance within a parallel system may never be achieved as there is always the possibility of the complete failure of all nodes. Checkpointing is one useful mechanism to reduce the cost of system failure by allowing the computation to continue from the last known valid state. This paper has shown that fault tolerance within the mobile parallel domain is far more complex. This is due to the numerous possibilities for failure inherent with a wireless communications medium. In the case of Bluetooth these additional failure cases include communications errors, devices moving in and out of the Bluetooth range, and nodes being terminated by user intervention.

One can clearly see that fault tolerance by the use of checkpointing within the mobile parallel domain is more than viable. This is achievable although at a slightly higher cost due to the higher possibility of failure than that of the dedicated systems for high end cabled parallel clusters.

5 Acknowledgements

The development of this body of work was funded under the “Irish Research Council for Science, Engineering and Technology” funded by the “National Development Plan”.

References

[1] A. Agbaria and R. Friedman, “Starfish: Fault-Tolerant Dynamic MPI Programs on Clusters of Workstations”, *The Eighth International Symposium on High Performance Distributed Computing*, pp. 167–176, 1999.

[2] ARM, “Arm introduces industry’s fastest processor for low-power mobile and consumer applications”, `http://www.arm.com/news/10548.html`, October 2005.

[3] ARM, “Arm neon technology fuels consumer electronics growth with next-generation mobile multimedia acceleration”, `http://www.arm.com/news/6540.html`, October 2005.

[4] A. Baker, “Mini Review - Enhancements in Nokia 6680,” `http://www.i-symbian.com/forum/images/articles/43/Mini_Review-Nokia_6680_Enhancements.pdf`, 2005.

[5] Bluetooth SIG, “Annex A (Normative): Timers and Constraints”, *Bluetooth Specification version 1.1*, 2001.

[6] G. Bosilca, A. Bouteiller, F. Cappello, S. Djilali, G. Fedak, C. Germain, T. Herault, P. Lemarinier, O. Lodygensky, F. Magniette, V. Neri, and A. Selikhov, “MPICHV: Toward a Scalable Fault Tolerant MPI for Volatile Nodes”, *Supercomputing, ACM/IEEE Conference*, pp. 29–29, 2002.

[7] D. C. Doolan, S. Tabirca, L. T. Yang, “Mobile Parallel Computing”, *5th International Symposium on Parallel and Distributed Computing (ISPDC06)*, pp. 161–167, 2006.

[8] G. E. Fagg, J. J. Dongarra, “FT-MPI: Fault Tolerant MPI, Supporting Dynamic Applications in a Dynamic World”, *Lecture Notes in Computer Science*, vol. 1908, pp. 346–353, 2000.

[9] W. Gropp, E. Lusk, “Fault Tolerance in MPI Programs”, *Cluster Computing and Grid Systems Conference*, `http://www-unix.mcs.anl.gov/~gropp/bib/papers/2002/mpi-fault.pdf`, December 2002.

[10] R. Jaques, “One billion mobile phones shipped in 2006”, `http://www.itweek.co.uk/vnunet/news/2173516/billion-mobile-phones-shipped`.

[11] S. Louca, N. Neophytou, A. Lachanas, and P. Evripidou, “Mpi-ft: Portable fault tolerance scheme for mpi”, *Parallel Processing Letters*, vol. 10, no. 4, pp. 371–382, 2000.

[12] Nokia, “Nokia n70 technical specs”, `http://www.forum.nokia.com/main/0,,018-2578,00.html?model=N70`, April 2005.

[13] F. Pilato, “Arm reveals 1ghz mobile phone processors”, `http://www.mobilemag.com/content/100/102/C4788/`, October 2005.

[14] D. Robinson, “Arm chips to power 1ghz mobiles”, `http://www.vnunet.com/itweek/news/2143741/arm-chips-power-1ghz-mobiles`, October 2005.

[15] G. Stellner, "CoCheck: Checkpointing and Process Migration for MPI", *Parallel Processing Symposium, Honolulu, Hawaii* , pp. 526–531, 1996.

[16] S. Shankland, "Sun starts bidding adieu to mobile-specific Java", *Cnet News*, `http://www.news.com/8301-13580_3-9800679-39.html`.

[17] T. Tannenbaum and M. Litskow, "Checkpoint and Migration of Unix Processes in the Condor Distributed Processing System", *Dr. Dobbs Journal*, Vol. 227, pp. 40–48, 1995.

Daniel C. Doolan, Sabin Tabirca
University College Cork
Department of Computer Science
College Road, Cork, Ireland
E-mail: {d.doolan, s.tabirca} @cs.ucc.ie

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 271-276

Adaptive Neuro-Fuzzy Inference System for Mid Term Prognostic Error Stabilization

Otilia Dragomir, Rafael Gouriveau, Noureddine Zerhouni

Abstract: The high costs in maintaining complex equipments make necessary to enhance maintenance support systems and industrial and research communities take a growing interest in the "prognostic process". However, this activity is still not well bounded and real prognostic systems are scarce. Thus, the general purpose of the paper is to explore the way of performing failure prognostics so that manager can act consequently. The prognostic process is discussed from different points of view (concept, metrics, approaches and tools) in order to point out the pragmatic challenges of this activity. Assuming that maintenance decisions follow from a prediction step, the stabilization of mid term prediction errors appears to be essential. For that purpose a neuro-fuzzy predictor based on the ANFIS model is proposed to perform prognostic.
Keywords: prognostic, neuro-fuzzy system, ANFIS, error of prediction.

1 Introduction

Maintenance activity combines different methods, tools and techniques to reduce costs while increasing availability, reliability and security of equipments. That said, maintenance is far away to be only an industrial area of interest and researchers also show a growing attention to this thematic.
The initial maintenance framework was delimited by the necessity of "perceiving" phenomena, next, of "understanding" them, and finally, of "acting" consequently. However, rather than understanding a phenomenon which has just appeared like a failure, it seems convenient to "anticipate" it's manifestation in order to act consequently and resort to protective actions. This is what could be defined as the "prognostic process". Prognostic is nowadays recognized as a key feature in maintenance strategies. However, real prognostic systems are scarce in industry and that can be explained from different aspects. Firstly, prognostic still is not a stabilized concept: there is no consensual way of understanding it which makes harder the definition of tools to support it in real applications. Secondly, many approaches for prediction exist whose applicability is highly dependent of the available knowledge on the monitored system. Thirdly, the vagueness of prognostic process definition impedes to point out the inherent challenges for scientists. Thus, the purpose of this paper is to analyze and discuss the prognostic process from different points of view, and to propose a way of handling failure prognostics so that practitioners can act consequently.
The paper is organized in two parts. First, the prognostic framework is delimited, starting with the prognostic definition, metrics, approaches and tools. Considering prognostic as the association of a prediction and an evaluation steps, a classification of prognostic metrics is given. The whole aims at giving a frame to perform (and develop) real prognostic systems. In the second section, the use of neuro-fuzzy predictor for prognostic purpose is briefly justified and an illustration based on the adaptive neuro-fuzzy inference system is given. The proposed system performs "good" prediction and is built in order to reach the stabilization of mid term prognostic errors.

2 Performing prognostic: concept, measures and tools

2.1 Prognostic concept

Although there are some divergences in literature (see [2]), prognostic can be defined as proposed by the International Organization for Standardization: "prognostic is the estimation of time to failure and risk for one or more existing and future failure modes" [4]. In this acceptation, prognostic is also called the "prediction of a system's lifetime" as it is a process whose objective is to predict the remaining useful life (RUL) before a failure occurs given the current machine condition and past operation profile [6]. Two salient characteristics of prognostic can be pointed out [2]. 1) Prognostic is mostly assimilated to a prediction process (a future situation must be caught). 2) Prognostic is grounded on the failure notion, which implies that it is associated with a degree of acceptability (the predicted situation must be assessed with regard to a referential). Therby, at a prediction level, a prognostic system should be able to determine the future state of equipment as closely as possible to the future

real state. Also, the control of the performance of prediction is the premise of a good global prognostic system. At an evaluation level, the predicted situation should be evaluated regarding the reference levels, RUL, confidence, accuracy, etc., which implies the definition of prognostic measures.

2.2 Prognostic metrics

There is no general agreement as to an appropriate and acceptable set of metrics that can be employed effectively in prognostic applications, and researchers and CBM practitioners are still working on this [7].

1) The main objective of prognostic is to provide the efficient information that enable the underlying decision process, i.e., the choice of maintenance actions. Thus, a first set of metrics are those that quantify the risks incured by the monitored system. This kind of metrics can be called the **prognostic measures**.
As mentioned earlier, the main prognostic measure pursued is the predicted time to failure (**TTF**), also called the remaining useful life (**RUL**). In addition, a **confidence** measure can be built to indicate the degree of certitude of the future predicted failure time. By extension, and considering that practitioners can be interested on assessing the system with regard to any performance limit, RUL and confidence can be generalized: in Fig. 1(a), **TTxx** refers to the remaining time to overpass the performance limit Perf/xx, and **Conf/xxT** is the confidence with which can be taken the asset TTxx $>$ T.

2) Assuming that prognostic is in essence an uncertain process, it is useful to be able to judge from its "quality" in order to imagine more suitable actions. In this way, different indicators can be constructed: the **prognostic system performance measures**.
The **timeliness** of the predicted time to failure (TTF) is the relative position of the probability density function (pdf) of the prediction model along the time axis with respect to the occurrence of the failure event. This measure evolves as more data are available and reveals the expected time to perform preventive actions [7] (see Fig. 1(b)). According to [3], one needs to define two different boundaries for the maximum acceptable late and early predictions.
Accuracy measures the closeness of the predicted value to the actual value. It has an exponential form and is as higher as the error between the predicted value of TTF and the real one is smaller. **Precision** measure reveals how close predictions are grouped or clustered together and is a measure of the narrowness of the interval in which the remaining life falls. Precision follows from the variance of the predicted results for many experiments. The complementarity of accuracy and precision is illustrated in Fig. 1(c).

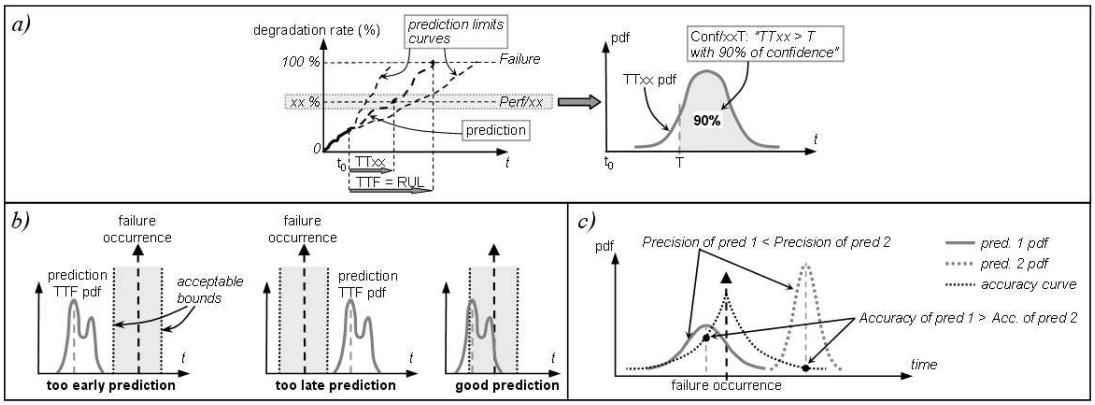


Figure 1: Metrics - (a) RUL, TTxx and confidence, (b) timeliness, (c) accuracy and precision

2.3 Prognostic approaches

Various prognostic approaches have been developed ranging in fidelity from simple historical failure rate models to high-fidelity physics-based models [7]. Similarly to diagnosis, prognostic methods can be associated with one of the following two approaches, namely model-based and data-driven.

1) Model-based methods assume that an accurate mathematical model for the analyzed system can be constructed. The main advantage of these approaches is their ability to incorporate physical understanding of monitored system. Moreover, if the understanding of the system degradation improves, the model can be adapted to increase its accuracy and to address subtle performance problems. But, this closed relation with a mathematical model may also be a strong weakness: it can be difficult, even impossible to catch the system's behavior.

2) Data-driven approaches use real data to track, approximate and forecast features revealing the degradation of components; in many applications, measured input/output data is the major source for a deeper understanding of the system degradation. Data-driven approaches can be divided into two categories: statistical techniques (multivariate statistical methods, linear and quadratic discriminators, etc.), and artificial intelligence (AI) techniques (neural networks, fuzzy systems, etc.). The strength of data-driven techniques is their ability to transform high-dimensional noisy data into lower dimensional information for decisions. In practice however, data-driven approaches are highly-dependent on the quantity and quality of operational data.

3 Building a neuro-fuzzy prognostic tool

3.1 ANFIS model: an adequate prediction tool

Real systems are complex and their behavior is often non linear, non stationary. These considerations make harder a modeling step, even impossible. Yet, a prediction computational tool must deal with it. Moreover, monitoring systems have evolved and it is now quite easy to online gather data. According to all this, data-driven approaches have been increasingly applied to machine prognostic. More precisely, works have been led to develop systems that can perform nonlinear modeling without a priori knowledge, and that are able to learn complex relationships among "inputs and outputs" (universal approximators). Indeed, artificial neural networks (ANNs) have been used to support the prediction process. Recent works focus on the interest of hybrid systems: many investigations aim at overcoming the major ANNs drawback (lack of knowledge explanation) while preserving their learning capability. In this way, neuro-fuzzy systems are well adapted. More precisely, first order Tagaki-Sugeno (TS) fuzzy models have shown improved performances over ANNs and conventional approaches [8]. Thereby, they can perform the degradation modeling step of prognostic.
A particular architecture of TS neuro-fuzzy systems is that of the adaptive neuro-fuzzy inference system (ANFIS) [5]. ANFIS is an inference system in which the parameters associated with specifics memberships functions are computed using either a backpropagation gradient descent algorithm alone or in combination with a least squares method. Thanks to its structure and learning capability, ANFIS is fitted to predict irregular or non-periodic time series. However, when used for mid term predictions purpose, ANFIS can make large residual errors. Next part of the paper emphasizes on this aspect.

3.2 Mid term prognostic error stabilization with ANFIS predictor

Since the occurrence of a failure is in essence uncertain, a prognostic tool should be enable to make predictions with quite the same accuracy at short, mid and long terms. This is the purpose of this part (assuming that ANFIS is an adequate short term prediction tool).

1) Here, prognostic is considered as a prediction process based on the aggregation of, naturally, past and present states of the system, but also, of the known future ones. Indeed, it appears to be useful to take into account future actions like the modification of the mission profile due to some extern interventions or like the influence of a scheduled maintenance action. Consequently, the just-in-time-point (the time of failure when the life duration is [0%]) gets another dimension related to the starting point ([100%] of machine life duration) (see Fig. 2).

2) In most of the papers in which ANFIS is used as a prediction system, inputs are directly extracted from the data sets. Here, the Box-Jenkins furnace benchmark is used. There are originally 296 data samples {y(t), u(t)}, from t=1 to t=296. From the real process, CO2 concentration is considered as the output of the model y(t), and gas flow rate as the input u(t). In order to predict y(t) based on y(t-1), y(t-2), y(t-3), y(t-4), u(t-1), u(t-2), u(t-3), u(t-4), u(t-5), u(t-6), the number of effective data points becomes 290. A selection method must be used because all ten input variables generate to many rules and parameters to be updated on the learning phase: it would make the training data insufficient and would obviously increase the computing time.

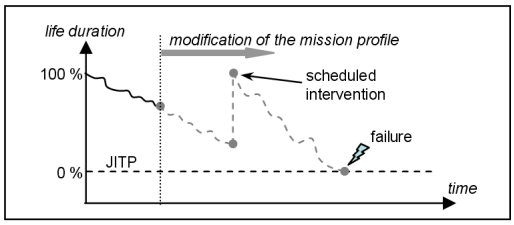


Figure 2: Influence of scheduled maintenance actions on the prediction process

3) The choice of an error measure to compare prediction methods has been much discussed (see for example [1]). In any case, error measures are only intended as summaries for the error distribution. This distribution is usually expected to be a normal white noise in a forecasting problem, but it probably is not so in a complex problem like load forecasting. The ANFIS architecture proposed by Jang with two and three selected inputs has satisfactory results for short term predictions. For mid and long terms ones, the obtained errors increase, which affects the prognostic performances (Fig. 3).

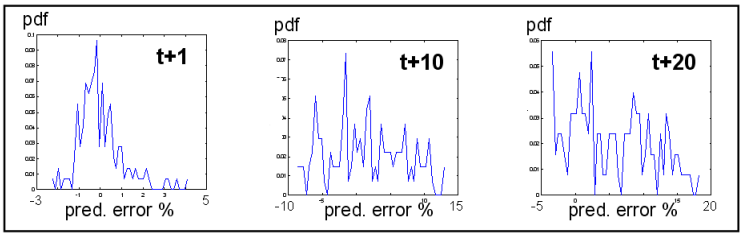


Figure 3: Error of prediction pdf - ANFIS model with two inputs

4) A first way to stabilize the error of prediction is to construct serial architectures of ANFIS models. Indeed, linking two ANFIS enable to take over the growth tendency of error since the second one learns the error of the first one. The effect of this modification is observed on the error values (see Fig. 4). The maximum measured error decreases significantly and becomes satisfactory for mid term predictions as well as the prediction spread does.

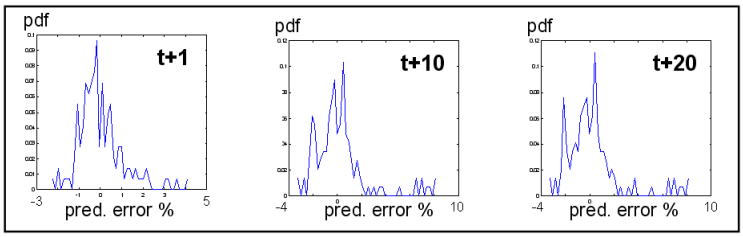


Figure 4: Error of prediction pdf - serial architecture of two ANFIS models with two inputs

5) The known future solicitations of the system (mission profile) can then be injected as input information for the second ANFIS module in the serial architecture. The effect of taking into account the “future” for predictions has also influence over error: the error spreading is significantly reduced and the confidence of the prediction process increases thereby consequently. Mid term predictions reflect the improved quality of the approach (Fig. 5).

4 Conclusion and work in progress

In this article, prognostic is presented as the association of a prediction and an evaluation process. Many approaches to support this activity exist, whose applicability and performance must be assessed to develop a prognostic tool. For that purpose, various prognostic metrics exist and are discussed in the paper. All of them are based on the evaluation of the prediction error spreading. Following that, the prediction step of prognostic appears

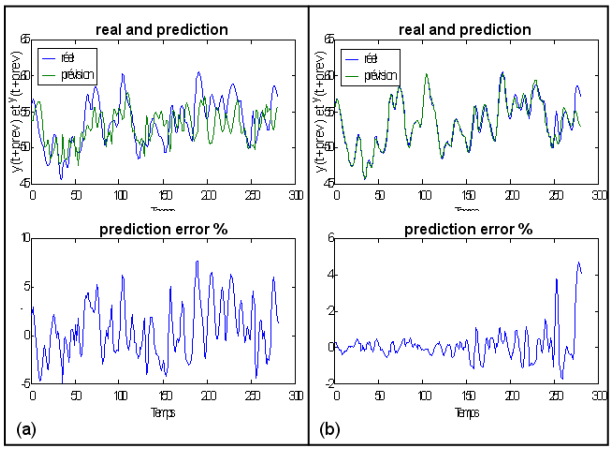


Figure 5: Error of prediction (%) for a serial architecture without (a) and with (b) future known solicitations injection

to be a critical one and controlling the performances of predictions is the premise for a good global prognostic system. According to all this, the interest of neuro-fuzzy system is pointed out and a new ANFIS architecture is proposed in order to ensure a certain stability of errors for mid term predictions. The system is based on a serial architecture of various ANFIS models, in order to overcome the local prediction errors (by learning them) and to inject the future scheduled maintenance actions. The obtained results are satisfactory from the industrials point of view since confidence on predictions increases.
The work is still in progress and the developments are at present extended in four principal ways. First, the definition of new loss functions in the learning phase is studied. Secondly, the application of ANNs and NFs as tools for a global prognostic is been investigated. Thirdly, the interpretability of the obtained predictive system is been looked in a closely manner. Finally, the implementation of the studied framework is in progress at a French industrial partner for the monitoring of high speed trains motors.

References

[1] J.G. De Gooijer, R.J. Hyndman, "25 years of time series forecasting", *International Journal of Forecasting*, Vol. 22, pp. 423-473, 2006.

[2] 0. Dragomir, R. Gouriveau, N. Zerhouni, "Framework for a distributed and hybrid prognostic system", *In 4th IFAC Conference on Management and Control of Production and Logistics*, 2007.

[3] K. Goebel, P. Bonissone, "Prognostic information fusion for constant load systems", *In: Proceedings of 7th annual Conference on Fusion*, Vol. 2, pp. 1247-1255, 2005.

[4] ISO 13381-1, *Condition monitoring and diagnostics of machines - prognostics - Part1: General guidelines*, International Standard, ISO, 2004.

[5] J-S.R. Jang and C. Sun, "Neuro-fuzzy modeling and control", *In: IEEE Proceedings*, Vol. 83, pp. 378-406, 1995.

[6] A.K.S. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance", *Mech. Syst. and Sig. Process.*, Vol. 20, pp. 1483-1510, 2006.

[7] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess and B. Wu, *Intelligent Fault Diagnosis and Prognostic for Engineering System* , New-Jersey, Hoboken: Wiley and Sons, 2006.

[8] W-Q. Wang, M.F. Goldnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems", *Mech. Syst. and Sig. Process.*, Vol. 18, pp. 813-831, 2004.

Otilia Dragomir(1), Rafael Gouriveau(2), Noureddine Zerhouni(2)
(1) Valahia University of Târgoviste, Electrical Engineering Faculty,
Automation and Information Department,

Unirii Avenue 18-20, Târgoviste, Romania
(2) FEMTO-ST Institute, CNRS - UFC / ENSMM / UTBM,
Automatic Control and Micro-Mechatronic Systems Department,
24 rue Alain Savary, 25000 Besançon, France
E-mail: drg_otilia@yahoo.com, rafael.gouriveau@ens2m.fr, noureddine.zerhouni@ens2m.fr

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 277-281

Justifying GIS Modeling Uncertainty: A Practical Approach

Gabriela Droj

Abstract: The usage of Geographic Information Systems (GIS) has been rapidly increased and it became the main tool for analyzing spatial data in many engineering applications and decision making activities. These applications and activities require more and more the three-dimensional reference space (surface), known as Digital Terrain Model (DTM). Obviously, in order to take the correct decisions it is necessary to have high quality with known inaccuracy DTMs. The objective of this article is to compare a number of different algorithms involved in producing a DTM for GIS modeling applications, as well as to establish a representative set of parameters necessary to quantify the uncertainty of these DTMs.
Keywords: GIS, Spatial Data Accuracy, Data Quality, Modeling Uncertainty.

1 Introduction

In the recent years, the usage of Geographic Information Systems (GIS) has been rapidly increased and it became the main tool for analyzing spatial data in a number of scientific fields of real world activities. The Geographic (or Geographical) Informational Systems are used not only in assisting daily work, but as well as a decision support tool especially for planning (urban planning, investment planning, infrastructure planning, economic development, taxation, SWOT analysis, etc.), resource management and fiscal impact [2,8,11].
GIS are also used in critical systems like civil protection which need high performance and fast computing time [4]. These applications require more and more the three-dimensional reference space (actually a 2.5D surface), although temporary aspects demand for four dimensions? including time as the additional (temporal) reference parameter [12,13].
Using GIS technology, contemporary maps have taken radical new forms of display beyond the usual 2D planimetric paper map. Today, it is expected to be able to drape spatial information on a 3D view of the terrain. The 3D view of the terrain is called Digital Terrain Model (DTM) or Digital Elevation Model (DEM) 3 .
The digital terrain models are frequently used to take important decisions like to answer hydrological questions concerning flooding. In order to judge these decisions the quality of DTM must be known. The quality of DTM is, unfortunately, rarely checked. While the development of GIS advances, DTM research has so far been neglected.
The objectives of this paper are: (a) to compare the different algorithms involved in producing a DTM, and (b) to establish a representative set of parameters necessary to quantify the uncertainty of DTMs.

2 DTM for GIS Modeling Applications

A DTM is a digital representation of ground surface, topography or terrain. It is also known as Digital Elevation Model (DEM). The DTM can be represented as a raster (a grid of squares), as contour lines or as a triangular irregular network (TIN) [5,7,10] .
In GIS applications and according to the relative bibliography, the following two methodogies are frequently used for DTM GIS 2.5D modeling: (i) the cartographic interpolation-based digitizing method, and (ii) the image-based automatic measurements method 9 .

1. The cartographic interpolation-based digitizing method is widely used because topographic maps are usually available. The input data form the basis for the computation of the DTM, consisting of points. The computation itself consists in spatial interpolation algorithms.

2. The image-based automatic measurements method is based on close-range photogrammetry or airborne laser scanning and outputs bulk points with a high density. The DTM for GIS 2.5D modeling applications is realized in the post processing phase usually by creating the TIN or by interpolation

2.1 Interpolation-based DTMs

The current research and industrial projects in GIS require higher standards for the accuracy of spatial data. The data in geographical informational systems (GIS) are usually collected as points, where a point is considered

a triplet (x,y,z), where x and y are the coordinates of the point and z is the specific information. This specific information can be for example: the altitude level in the point (x, y), the quantity of precipitations, the level of pollution, type of soil, socio-economic parameters etc.. The mapping and spatial analysis often requires converting this type of field measurements into continuous space. Interpolation is one of the frequently used methods to transform field data, especially the point samples into a continuous form such as grid or raster data formats.
There are several interpolation methods frequently used in GIS. The following eight widely used methods are compared and studied in this paper. These methods are: Inverse distance weighted (IDW), Spline Biquadratic interpolation, Spline Bicubic interpolation, B-spline interpolation, Nearest neighbours - Voronoi diagrams, Delaunay triangulation, Quadratic Shepard interpolation and Kriging interpolation 1,6,14 .

2.2 Image-based DTMs

The fastest way of obtaining the digital elevation model is by using remote sense data. In the case of collecting data with close-range photogrammetry or airborne laser scan the result consist in a high density of points with three coordinates (x,y,z). By computing the TIN model of these points we can obtain fast a DTM. The quality of the DTM is dependent by the quality of the image, especially by the parameter z which has always a lower accuracy that the pair (x, y).
By using this method we can obtain fast a DTM, this method is used especially in civil protection in case of disaster. This method it's not recommended to be used for urban areas because the value of z is equal with the sum between the altitude and the high of the object.

3 Quantifying Uncertancy in GIS Modeling

Measurement of errors for the results is often impossible because the true value for every geographic feature or phenomenon represented in this geographic data set is rarely determinable. Therefore the uncertainty, instead of error, should be used to describe the quality of an interpolation result. Quantifying uncertainty in these cases requires comparison of the known data with the created surface 15].
To analyze the pattern of deviation between two sets of elevation data, conventional ways are to yield statistical expressions of the accuracy, such as the root mean square error, standard deviation, mean, variance, coefficient of variation. In fact, all statistical measures that are effective for describing a frequency distribution, including central tendency and dispersion measures, may be used, as long as various assumptions for specific methods are satisfied 9,15 .
For the evaluation of DTM the most widely used measure, usually the only one, is the well known Root Mean Square Error (RMSE). Actually, it measures the dispersion of the frequency distribution of deviations between the original elevation data and the DTM data, mathematically expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(z_{d,i} - z_{r,i}\right)^2}$$

where: $Z_{d,i}$ is the ith elevation value measured on the DTM surface;
$Z_{r,i}$ is the corresponding original elevation;
n is the number of elevation points checked.

4 DTM Quality Evaluation: Experiments with Real-World Data

DTMs are the most popular results of interpolation. In the following we will test different methods and algorithms for creation of DTM in order to establish a minimum set of parameters to compare and evaluate the quality of the resulted data [8,11].
To test and compare the methods with real data we have selected an area from north hills of Oradea municipality.
For the first DTM we used photogrammetric measurement of spot elevations from orthorectified airborne image of the area. The TIN of the area was generated using ArcGIS Desktop 9.1. In the pictures below the created 3D model is presented (Fig. 1).

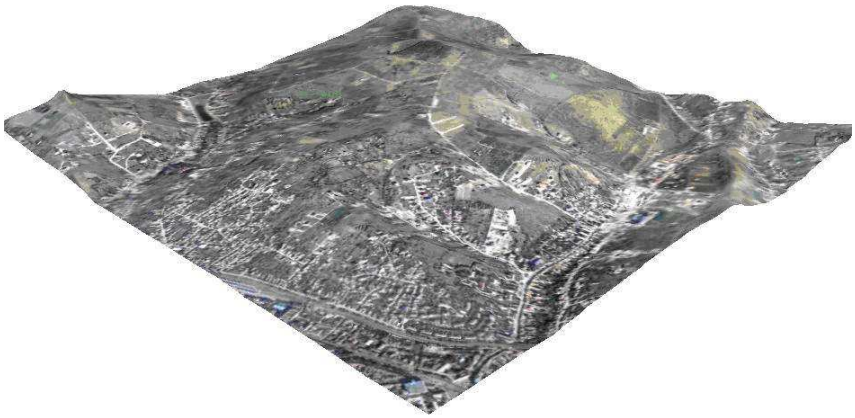


Figure 1: The Oradea 3D Model

This model will consider the reference for testing and evaluating the most popular eight interpolation algorithms used for DTM creation (please see Section 2).
The first step was to create a regular grid with the step of 500 m, for a total of 60 points. On this set of points we tested the algorithms specified before. The analyses of the results were made by direct observation, with visual comparisons of the models and by using statistical parameters. The visual comparisons show a high similarity with the reference for the Delaunay triangulation and Shepard interpolation.
For the statistical evaluation we determined the median absolute deviation, the standard deviation, the coefficient of variation and RMSE. In Table 1 the statistics for all the created surfaces are presented (Table 1).
By comparing the statistic evaluation and the real world situation we notice an inconsistence. The highest statistical accuracy is given by lowest value for RMSE, in this case this is the B-cubic interpolation.

Methods	Time	Min	Max	Mean	Median absolute deviation	Standard deviation	Coefficient of variation	RMSE
Initial data		125	180	245	25	30.9	0.172	
IDW	0.3	125	180	245	11.92	20.05	0.110	182.66
Bi-quadratic	0.1	129	181.7	222.9	13.96	20.17	0.110	183.11
Bi-cubic	0.1	161.7	179.5	197.3	6.99	8.39	0.46	179.75
B-spline	0.8	110	177.9	246.4	19.2	27.73	0.151	184.96
Voronoi	0.08	125	180	245	25	29.69	0.16	184.77
Delaunay	0.03	125	179	244.6	16.11	24.6	0.135	183.9
Shepard	0.3	100	173	245	20.3	27.6	0.152	183
Kriging	0.14	125	180	245	20.6	28.67	0.160	180.82

Table 1: The statistics for all the created surfaces with 60 known points

In the second step we have created a regular grid with the step of 250 m, for a total of 121 points. On this set of points we tested the same algorithms as before. The visual comparisons show a high similarity, with the reference, for the Shepard interpolation, Kriging interpolation and Delaunay triangulation.
The statistical data on the surfaces created show a decrease of the value for all the parameters, as it can be seen in the Table 2.

In the second case, as in the first one, regarding the statistical parameters the optimal surface is generated with Be-quadratic interpolation but the direct observation shows a higher similarity with the reference for the Shepard interpolation, Kriging interpolation and Delaunay triangulation.
The results obtained in the both cases show that the statistical analyses are inconsistent and insufficient, if it's done

Methods	Time	Min	Max	Mean	Median absolute deviation	Standard deviation	Coefficient of variation	RMSE
Initial data		125	180	245	25	30.2	0.166	
IDW	0.11	125.5	176.09	242.87	14.30	21.82	0.121	181.11
Bi-quadratic	0.3	122.4	176	227	17.89	22.91	0.129	179
Bi-cubic	0.3	164.8	180	195	6.06	7.2	0.04	180.1
B-spline	3.92	134.7	172.7	238.85	20.5	29	0.163	179.9
Voronoi	0.11	125	175	245	25	30	0.16	181.01
Delaunay	0.11	125	177	244.8	20.3	27.6	0.152	183
Shepard	0.14	125	177.84	245	19.09	27.1	0.148	185
Kriging	0.08	125	180	245	17.17	25.63	0.140	184.26

Table 2: The statistics for all the created surfaces with 121 known points

on the same data as the interpolation algorithms. This suggests the need of testing the quality of the surfaces with an independent set of data, which were not considered in the interpolation.
In the following we will evaluate the surfaces generated by using a random set of points for which was determined the real value of the altitude. For this independent random set we have determined the following statistical parameters: variation, median absolute deviation, standard deviation and the root mean square error. The determined values are presented in the table 3.

Methods	Variation 250	Variation 500	Median absolute deviation 250	Median absolute deviation 500	Standard deviation 250	Standard deviation 500	RMSE 250	RMSE 500
IDW	231	167	11.6	10.17	15.21	12.29	13.48	10.93
Bi-quadratic	390	426	16.17	16.56	19.75	20.64	15.35	13.84
Bi-cubic	947	1044	26.53	26	30.77	32	19.14	17.14
B-spline	166.2	145	9.67	9.9	12.89	12	12.54	10.37
Voronoi	246	94.5	11.38	7.3	15.71	9.75	13.79	9.27
Delaunay	218	36.63	10.85	3.41	14.78	6.05	13.34	7.99
Shepard	164	71.78	9.56	6.40	12.81	8.47	11.97	8.49
Kriging	160.06	78.97	9.74	6.81	12.65	8.88	11.75	8.6

Table 3: The statistics for all the created surfaces, determined for a set of random control points

5 Summary and Conclusions

By evaluating the statistical data we noticed that the most accurate surfaces are generated, for the first case (grid of 500 m), by Kriging, Shepard and B-spline algorithms and for the second case (grid of 250m) by Delauney triangulation followed by Shepard and Kriging interpolation. If we evaluate all the statistical data (presented in the all three tables), we notice that the Delauney triangulation is representing the optimal method. Similar results can be obtained by Kriging and Shepard interpolation. Even these methods are sometimes more efficient than the Delauney triangulation. Nevertheless, the Delauney algorithm is recommended because it needs less computing time and it is not changing the original values of the points. The B-spline algorithm also gives a good result but in this case the computing time is much higher and it is smoothing the surface, fact which is making this method inadequate for surfaces with a high altitude difference. The first conclusion, which is pointed up by the values presented in Table 3, is that all the statistical indices are pointing out a similar order of accuracy. Another conclusion is that the statistical evaluation of the random set of data illustrated a higher accuracy for a higher number of initial known points. The last conclusion is that, for an accurate evaluation, is necessary to evaluate the surface by using an independent set of data. In this case the RMSE mirrors the quality of the surface but for a

correct evaluation we recommend to analyze at least one more statistical index like standard deviation or Median absolute deviation

References

[1] M. R. Asim, M. Ghulam, K. Brodlie, "Constrained Visualization of 2D Positive Data using Modified Quadratic Shepard Method," *WSCG'2004*, Plzen, Czech Republic.

[2] G. Benwell "A Land Speed Record? Some Management Issues to Address," *Int'l Conference on Managing GIS for success*, Melbourne, Australia, pp. 70-75, ISBN: 0 7325 1359 6, 1996.

[3] J. K. Berry and Associates "The GIS Primer - An Introduction to Geographic Information Systems," *http://www.innovativegis.com/basis/*, May 2006.

[4] Tai On Chan, Ian Williamson "A Holistic, Cost-Benefit Approach to Justifying Organization-Wide GIS," *Int'l Conference on Managing GIS for success*, Melbourne, Australia, pp. 27-35, ISBN: 0 7325 1359 6, 1996.

[5] Du Chongjiang "An Interpolation Method for Grid-Based Terrain Modeling," *The Computer Journal*, Vol. 39, No. 10, 1996.

[6] ESRI. Environmental Science Research Institute "Arc/Info 8.0.1," *ESRI*, Inc. 1999.

[7] L. De Floriani, E. Puppo, P. Magillo "Application of Computational Geometry to Geographic Informational Systems," *Handbook of Computational Geometry*, 1999 Elsevier Science, pp. 333-388.

[8] Gary J. Hunter " Management Issues in GIS: Accuracy and Data Quality," *Int'l Conference on Managing GIS for success*, Melbourne, Australia, pp. 95-101, ISBN: 0 7325 1359 6, 1996

[9] W. Karel, N. Pfeifer, C. Briese " DTM Quality Assessment," *ISPRS Technical Commission II Symposium* , 2006 XXXVI/2, pp. 7-12.

[10] M. van Kreveld " Algorithms for Triangulated Terrains," *Conference on Current Trends in Theory and Practice of Informatics*, 1997.

[11] A. D. Styliadis " Risk Management for GIS Projects - Lessons Learnt," *Int'l Conference on Managing GIS for success*, Melbourne, Australia, pp. 102-111, ISBN: 0 7325 1359 6, 1996.

[12] A. D. Styliadis, M.Gr. Vassilakopoulos "A Spatio-temporal Geometric-based model for Digital Documentation of Historical Living Systems," *Elsevier Journal Information and Management*, Vol. 42, No. 2, pp. 349-359, 2005.

[13] A. D. Styliadis, "GIS: From Layers to Features," *Australian Magazine for CAD and MicroStation*, Pen & Brush Publishers Vol. 6, No. 3, pp. 18-29, 1996.

[14] R. T. Trâmbiţaş "Analiză numerică.Note de curs," *http://math.ubbcluj.ro/ tradu/narom.html*, 2003

[15] Qihao Weng, "Quantifying Uncertainty of Digital Elevation Model," *http://isu.indstate.edu/qweng,* December 2006.

Gabriela Droj
University of Oradea
Department Geodesy and Topography
City Hall of Oradea - Chief of GIS Department
E-mail: gaby@oradea.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 282-286

Evolutionary Programming in Disassembly Decision Making

Luminiţa Duţă, Florin Gheorghe Filip, Ciprian Popescu

Abstract: Disassembly retrieves components and materials from end-of-life products for remanufacturing, reuse and recycling. An essential criterion for a performing disassembly system is the benefit it brings, that is the revenue brought by the retrieved parts and material, decreased by the cost of their retrieval. A well balanced line will decrease the cost of disassembly operations. An evolutionary algorithm is used to deal with the multi-criteria optimization problem of the disassembly scheduling.
Keywords: process control, scheduling algorithms, genetic algorithms

1 Introduction

Disassembly of manufactured products induces both disassembly costs and revenues from the parts saved by the process. At planning stage a good trade-off has to be found that depends, both on the "depth" of the disassembly, and on the sequence of operations. The optimization of the ratio between gain and cost can be accomplished by using an appropriate distribution of the disassembly tasks on workstations. The optimization problem depends upon the structure of the disassembly system: if it is made up of a single workstation, the costs depend mainly upon the process duration. If the system is a line, the costs depend mainly upon the line balancing, all the more if it is highly manual. Another problem that occurs during a disassembly process is how deep the disassembly sequence must go so as to maximize the outcome of this process. In ([1, Duţă et al], [6, Duţă et al]) it was shown that an incomplete disassembly sequence can be more profitable than a complete one. Destructive and dismantling operations have to be taken into consideration, as well.

Hence, we have to deal with a multi-criteria optimization problem: maximizing the benefit it brings deciding how deep the disassembly sequence can go and minimizing the costs using an optimal scheduling along the line. A decision in a scheduling problem upon many criteria is a NP-hard to solve problem ([4, Filip]). Stochastic algorithms have already been used to fulfil a multi-criteria optimization problem in ([7, Minzu]).

In this paper we consider that the line structure was given and propose an algorithm which will allow finding a disassembly sequence and its assignment on workstations that optimizes a very simple function which integrates the income from the parts and the cycle time of the disassembly line.

2 The optimization problem

In this paper we address the case of disassembly lines where the cycle-time is not merely the sum of all operative and logistic times but it also depends strongly upon the line balancing. The objective is to find the most profitable disassembly sequence taking into account, on one hand - the end-of-life options for each part or subassembly of a given product, and on the other hand - the operational times for a given assignment of the tasks on the disassembly workstations. A cost function which combines both disassembly costs and revenues was proposed in ([5, Duţă et al]).

$$f = \frac{r}{t_{cy}} \tag{1}$$

where r is revenue associated to each disassembled part and t_{cy} is the cycle time.

The global revenue, r, is the sum of partial revenues obtained according to the end-of-life destinations of the disassembled parts. These partial revenues are established by experts after repeated disassembly processes.

$$r = \sum_k r_k \quad k = \overline{1..nc} \tag{2}$$

Where nc is the number of final components or the number of subassemblies obtained after the disassembly process.

The cycle time can be defined like the operational time of the slowest workstation on the line

$$t_{cy} = \max_{W_i} \sum_{j \in (tasks\ on\ W_i)} t_j \tag{3}$$

Where W_i is the i-th workstation, j is the index of the disassembly operation which requires the operational time t_j.

We make the following assumptions: a)the disassembly line is linear (flow-shop type), b) end-of-life revenues of the subassemblies are known, c) operational costs are included in the final incomes, d) the criterion of maximizing the outcome depending of the success rate of disassembly operations has been taken into consideration, e) the failure of the disassembly process is an event that can also occur since certain parts of the product could be deformed and impossible to be separated without destruction, f) the disassembly line works in a continuous flow regime.

Evaluating the function from (1) reveals the profit on a time unit, which is an important indicator for the productivity of the disassembly system. This function also takes into account the value of the cycle time obtained for a well-balanced line. The optimization can be made both for the manual and automatic disassembly lines.

3 Evolutionary programming

Genetic algorithms *(GA)* used in many research areas due to its capacity to reduce the combinatorial complexity of NP-complete problems. A genetic algorithm starts with a set of randomly generated possible solutions called *initial population*. Each member of a population is encoded as a *chromosome*. Chromosomes are represented by a combination of numbers or characters which contain information about the solution. A score named *fitness* is assigned to each chromosome based on the viability of the solution. Chromosomes with high scores are chosen as parents to create a new population. The objective is to obtain children with better scores. To avoid the uniformity of the population and to increase the space of research, at each step of creation, two processes may occur: *the crossover* and *the mutation*. Crossover combines the features of two or more parents into one child chromosome. Mutation generates a child similar with his parent with one or more genes altered. These operations ensure the diversity of the new generated population ([2, Goldenberg]).

Once established the initial population and defined the three types of operations, a genetic algorithm provides new members of population until a stop condition is fulfilled. Usually, this criterion is given by a maximal/minimal value of the objective function obtained after a number of iterations of the genetic algorithm.

4 Example

Consider the disassembly example of a Motorola radio set ([8, Salomonski and Zussman]). The corresponding Disassembly Petri Net(PN) is given in the **Figure 1**.

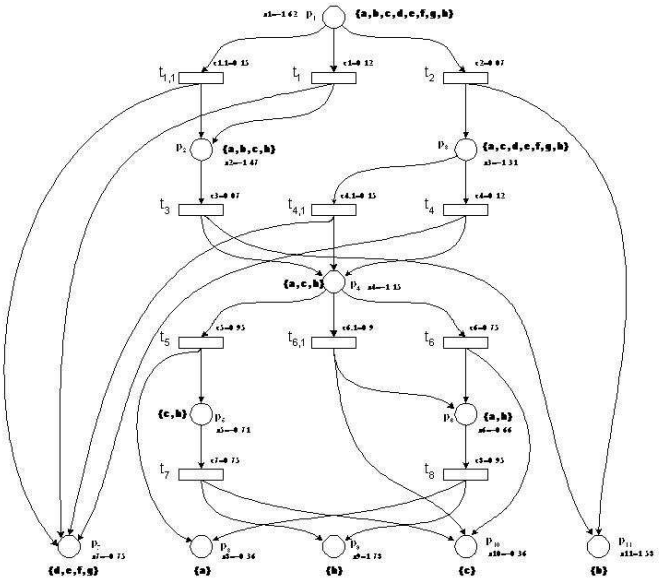


Figure 1: Disassembly Petri Net of a radio set

The disassembly process is represented four distinct possible sequences of transitions:
$\{t_1 \rightarrow t_3 \rightarrow t_5 \rightarrow t_7\}$, $\{t_1 \rightarrow t_3 \rightarrow t_6 \rightarrow t_8\}$, $\{t_2 \rightarrow t_4 \rightarrow t_5 \rightarrow t_7\}$ and $\{t_2 \rightarrow t_4 \rightarrow t_6 \rightarrow t_8\}$.

To take into account additional destructive disassembly methods, in the PN there are three alternative tasks represented by transitions $\{t_{1,1}, t_{4,1}, t_{6,1}\}$.

Assume that there are three workstations on the line. An alternative destructive task is done only on a so called "mixed" station that can perform both destructive and non-destructive operations. Thus, workstations 2 and 3 are considered mixed. In other words, when a task is moved from one station to another, the type of the operation is changed together with its operational time. We supposed that the tasks t_1 and t_4 can be performed in a non-destructive way on the first workstation and in a destructive way on the second one. For the task t_6 a destructive disassembly is done on the third station and a non-destructive disassembly is performed on the second workstation.

In accordance with PN of Figure 1 the correspondent operational times on the initial assignment of the tasks are:

$$T_i = \{0.92, 0.07, 0.07, 0.12, 0.95, 0.75, 0.75, 0.95\}$$

Taking into consideration the times for the alternative destructive operations the set above becomes:

$$T_f = \{0.95, 0.07, 0.07, 0.15, 0.95, 0.90, 0.75, 0.95\}$$

The final revenues are calculated after the method proposed in [1, Duta et al]) by using the data from PN of Figure 1. The corresponding sets of revenue values are:

$$R_i = \{-1.36, 0.27, 0.54, 0.54, 0.62, 0.67, 2.75, 2.75\}$$

$$R_f = \{-1.32, 0.27, 0.54, 0.50, 0.62, 0.60, 2.75, 2.75\}$$

The objective is to maximize the value of the function from the equation (1) by finding the sequence that maximize the final revenue and in the same time ensuring a well-balance of the disassembly line (e.g. minimizing the cycle time).

In our problem, a chromosome is represented by the possible tasks assignment matrix **S** whose entries are:

$$s_{ij} = \begin{cases} 1 & if\, the\, task\, j\, can\, be\, assigned\, to\, the\, workstation\, i \\ 0 & if\, the\, task\, j\, can't\, be\, assigned\, to\, the\, workstation\, i \end{cases}$$

$i = \overline{1..n}$ and $j = \overline{1..m}$ (n is the number of workstations and m - the number of disassembly operations).

A matrix **S** matrix can be the solution of our optimization problem only if it satisfies the operating constraints (*OC*):

1. The **non-divisibility constraint** that does not allow a task to be assigned to more than one station.

$$s_{ij} \in \{0, 1\} \tag{4}$$

2. The **assignment constraint** that requires that each task be assigned to *exactly* one station.

$$\sum_i s_{ij} = 1 \tag{5}$$

3. The **precedence constraint** that invokes technological order so that if task i is to be done before task j ($i < j$), then i cannot be assigned to a station downstream from task j

The steps of the genetic algorithm are:

Step 1 *Generating the initial population*. There are three workstations and eight operations (tasks). The matrix of the possible task assignment is:

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We generate 24 matrices **S** that fulfill the three constraints OC using the method presented in [5, Duta]. We randomly chose three of them as shown bellow:

$$S_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad S_2 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad S_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

The repartition of tasks on the workstations is represented in the **Figure 2**.

Step 2 *Evaluating by the value of the objective function*. The fitness coefficients f1 and f2 are defined by calculating the value of the objective function of the equation (1) for each complete disassembly sequence of the radio set shown in the Figure 1 (see Table 1).

Step 3 *Selection*. The selection of the individuals is made after their robustness so as to generate a more robust and healthy population. We can not utilize the roulette method of selection because it generates non valid individuals (that don't respect the three constraints specified before). Matrices S_1 and S_3 are the strongest.

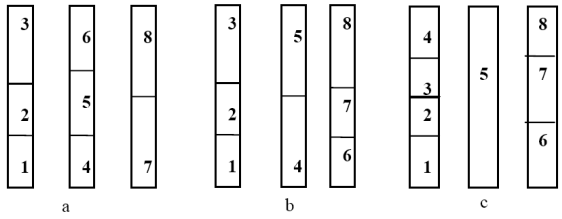


Figure 2: Assignment of tasks on the three stations for the initial population

Table 1: **Evaluation of the initial population**

	f1	*f2*	***fitness*** *(f1+f2)/2*	***strength*** *fitness %*	***average***
S1	0.5	0.6	0.55	33%	1
S2	0.5	0.43	0.46	27%	0.80
S3	0.75	0.6	0.675	40%	1.20
sum			1.685	100%	
average			0.5616	33%	

Table 2: **Evaluation of the new population**

	f1	*f2*	***fitness*** *(f1+f2)/2*	***strength*** *fitness %*	***average***
S3	0.75	0.6	0.675	31%	0.92
S5	0.95	0.75	0.85	38%	1.16
S6	0.75	0.6	0.675	31%	0.92
sum			2.20	100%	
average			0.73	33%	

Step 4 *Crossover*. If we have two matrices A and B of the same dimensions

$$A = \{a_{ij}\}, B = \{b_{ij}\}\ i = \overline{1..n},\ j = \overline{1..m}$$

The crossover operator $\oplus$ is defined as

$$a_{ij} \oplus b_{ij} = \begin{cases} 0\ if\ (a_{ij} = 0\ and\ b_{ij} = 0)\ or\ if\ (a_{ij} = 1\ and\ b_{ij} = 1) \\ 1\ if\ (a_{ij} = 1\ and\ b_{ij} = 0)\ or\ if\ (a_{ij} = 0\ and\ b_{ij} = 1) \end{cases}$$

We also define a special matrix called *mask matrix*

$$MSK = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Calculating $S_1 \oplus MSK$ and $S_3 \oplus MSK$ we obtained two chromosomes that have two genes crossed: a new chromosome and a copy of chromosome S_2:

Step 5 *Mutation*. A mutation can be made by the movement of one task between two neighboring stations. A new individual is obtained from S_4

Step 6 *Replacing the initial population*. Performances of the new population are represented in the table 2.

$$S_4 (= S_2) = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad S_5 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$S_5 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Step 7 *Iteration*. As a result the genetic algorithm is iterated until a stop condition is accomplished. After 20 iterations the maximal value of the objective function remains 0.95. So the optimal assignment of tasks is given by the matrix S_5. The value of the function from the equation (1) is given in euro/second. For the example presented the algorithm was implemented in the C++ language and executed on a AMD-Athlon processor at 1,8 Ghz.

5 Conclusions

A new computation method in the problem of the optimization of the disassembly sequences is proposed in this paper. The method used has the advantage that it takes into account the operational durations, as well as the profit achieved after a disassembly process from the valorization of the obtained components or subassemblies. In a balanced disassembly line, the cycle time has the lowest value so the operational costs are minimized. The algorithm does not optimize the balance of the disassembly line, but give a solution that improves this balance. When applying an evolutionary algorithm some undetectable solutions of the problem can be taken into account. However, using genetic algorithms implies a lot of information. The result is obviously faster obtained than using the backtracking method. In the disassembly process a local and fast solution for the optimal disassembly sequence is preferred to the complex and slower algorithms.

Acknowledgement. This work was partially supported from Grant CEEX 13-CEEX F03 of INCE-IEI funded by the National Authority for Scientific Research.

References

[1] L. Duţă, F.G. Filip, J.M. Henrioud, "Determination of the optimal disassembly sequence using decision trees", *Intelligent Assembly and Disassembly, ELSEVIER LTD*, pp. 43-48, 2003.

[2] D. E. Goldenberg, *Genetic Algorithms in Search, Optimisation and Machine Learning,* Adisson Wesley, 1989.

[3] *Informatique Industrielle,* vol. 29, no. 4-5, Edition Eyrolles, Paris, 1995.

[4] F.G. Filip, *Sisteme suport pentru decizii,* Ed. Tehnica, Bucuresti, 2004 (in Rom.).

[5] L. Duţă, F.G. Filip, J.M. Henrioud, "A method for dealing with multi-objective optimization problem of disassembly processes", *Proc. IEEE - ISATP03*, pp. 163-168, July, France, 2003.

[6] L. Duţă, J.M. Henrioud, I. Caciula, "A Real Time Solution to Control Disassembly Processes", *Proc. IFAC Conf., MCPL '07*, Sibiu, Sept. 2007.

[7] V. Minzu, J.M. Henrioud, "Stochastic algorithm for tasks assignments in single or mixed-model assembly lines", *APII-JESA*,vol. 32, no. 7-8, pp. 831-851, 1998.

[8] N. Salomonski, E. Zussman, "On-line Predictive Model for Disassembly Process Planning Adaptation", *Robotics and Computer Integrated Manufacturing*, vol. 15, pp. 211-220.

Luminiţa Duţă and Ciprian Popescu
Valahia State University, SSAI Dept.
Targoviste, Romania
E-mail: duta@valahia.ro, nicoc_ro@yahoo.com

Florin Gheorghe Filip
The Romanian Academy-INCE and ICI Bucharest
E-mail: ffilip@acad.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 287-292

An Application of Neuro-Fuzzy Modelling to Prediction of Some Incidence in an Electrical Energy Distribution Center

Simona Dziţac, Ioan Felea, Ioan Dziţac, Tiberiu Vesselenyi

Abstract: In this paper we will present the utilization of neuro - fuzzy models as prediction of some events, more exactly, realizing of some applications viewing the time intervals prediction in which incidents can appear in an electrical energy distribution system. It was realized the duration analyzes between two incidents, with the aim to estimate the frequency of the incidences in the future. The time intervals prediction where may appear incidences was realized for electric energy distribution center Oradea, the used language being MATLAB.
Keywords: neuro-fuzzy modelling, prediction, membership function

1 General considerations viewing the neuro-fuzzy modelling

In domain of reliability, the prediction of some values connected to halts, incidences or faults of distribution energy networks, is one from the major applications of the simulation methods. In this paper, the authors would find a method that may generate a model with prediction based on registering of the system parameter values. Such a model may be generated using methods of neuro-fuzzy modelling [2,3,4,5,7]. The case study was made by using the tables of "incidences", from the database of "quality indices of electric energy" [1, 6, 8].

The paper includes four sections. In the first section general considerations on applying neuro-fuzzy modeling to reliability analysis are presented. The second section is presenting the mathematical model involved in prediction of events. the third section deals with prediction of time intervals in which incidents can appear in an electrical energy distribution center of the city of Oradea. At the end conclusions and references are presented.

The underlying principle of fuzzy has ability to treat non precise and uncertain information, as neuronal networks may be identified by measuring of input and output signals of the process. The neuro-fuzzy networks combine the fuzzy reasoning with neural networks.

The basis idea of neuro-fuzzy methods is that the parameters of fuzzy system are computed through learning methods (rules of interference, membership functions) knowing data sets (inputs-outputs). The learning methods are the same as those that are used by neuronal networks.

In practice, de data sets aren't every time representative for the entire situation that may appear and for this reason must be allowed a certain level of error for which is accepted that the model operates satisfaction.

The testing is realized with data set that differs from the driving one (it is supposed unknown), and square average error obtained through the driving epoch is compared with those that are obtained from the checking set running. If, the two curves of error are convergent, the driving was correct made.

The membership function's adjusting parameter computing is realized with a gradient vector, which optimizes the correlation between neuro-fuzzy model and real system, specified by input-output data set. After the obtaining of gradient vector is applied an algorithm of optimization, defined by the real outputs differences square minimization and of those shaped. Adjusting of membership function parameters is made by a hybrid algorithm that uses the least small square method combined with back spreading errors [2, 4, 5, 7].

In MATLAB language for realization of neuro-fuzzy model is used the ANFIS module (adaptive neuro-fuzzy inference system).

2 Using neuro-fuzzy models for prediction of some events

In case of prediction of some values, basing on a serial values that represents a certain previous evolution, utilizing a predictive model based on inputs - outputs is non typical, because there is available only one data series.

Such a case represents the values of a time series, available until t moment, which values are utilized to find a next value of the series at t+P moment.

A method for such prediction is to consider a sequence of N values from the knowing interval, with step:

$$(x(t-(N-1)\Delta),\ldots,x(t-\Delta),x(t)) \tag{1}$$

For example, if it is chosen $\Delta = P$ for each value of the knowing series, at t moment, it may be defined a vector $w(t)$:

$$w(t) = [x(t-(N-1)P)x(t-(N-2)P)\ldots x(t)] \tag{2}$$

This vector $w(t)$ may be used as a set of input driving data for a neuro-fuzzy model. The output driving data set will be the prediction (or the estimation) $s(t)$:

$$s(t) = x(t+P) \tag{3}$$

Usually, the available data set is divided in two intervals: the first contains the driving data, as the second interval the checking data.

In driving interval, for each value will be calculate $w(t)$, as $w(t)$ sequence of values will be considered as the input data set, $s(t)$ sequence the output data set.

3 Application viewing the prediction of time intervals at which may appear incidences by a power system

In table of "incidences" from the database of "Quality indices of electric energy" [1,6,8], were registered the data at which appears incidences for each consumer.

In case of the above presented method of neuro-fuzzy, predictive model, the algorithm generates 16 rules with 104 parameters of membership function (considering membership functions as Gausian type). To rich satisfying results, it is important as the number of available driving data to be at least twice greater as the number of parameters of membership functions.

In case of the database with minimal set of 200 driving data were obtained only than, if the incidences were analyzed in centers of incidences from North Oradea Electrica Transilvania for a period of 6 years. Considering the preventive analyze for "centers", it were realized distinct studies for each center: Oradea, Stei and Alesd, the present paper represents analyzes only for center from Oradea.

In point of view of prediction, it was considered an interesting analyze of duration between two incidences. So it may be estimate as in the future what period of time will be exists between two incidences. So, it is possible to estimate in the future what periods of time will be exists between two consecutively incidences (or how dense will be in the future the incidences).

For realizing of analyzes, with MATLAB program information from the database (made in EXCEL), were decrypted with a program "bd_in4" special realized for this reason [1]. Then were sorted the interest "data" fields (data at which were take place the incidences) than were sorted the data on "centre" field with program "center selection" [1].

From the sorted information was computed the number of days between two incidences and it was made the neuro-fuzzy prediction for each center with the program of "F_predict_c1", "F_predict_c2" and "F_predict_c3" and presented in [1].

The anfis used parameters of membership function are: number of driving epoch = 150; error of driving limit = 150; initial value of optimizing step = 0,0001; decreasing rate of the optimizing step = 1,9; increasing rate of the optimizing step = 2,1.

For center Oradea, were defined the magnitudes of the data set and checking of 300 values. The results of the program [1], are presented in diagrams from figures 1 to 6.

4 Conclusions

Analyzing the diagram of the error comparing - figure 5 - it is remarked that the errors are convergent to a constant value and after a number of 150 epoch these remains at a constant value.

In figure 6, the prediction values are in interval of 300-600 numbers of incidences. For center in Oradea, the prediction values are considered satisfactory in 300-500 incidences. In these intervals, the prediction error is 1-2 days. It may be established very good predictions for center in Oradea.

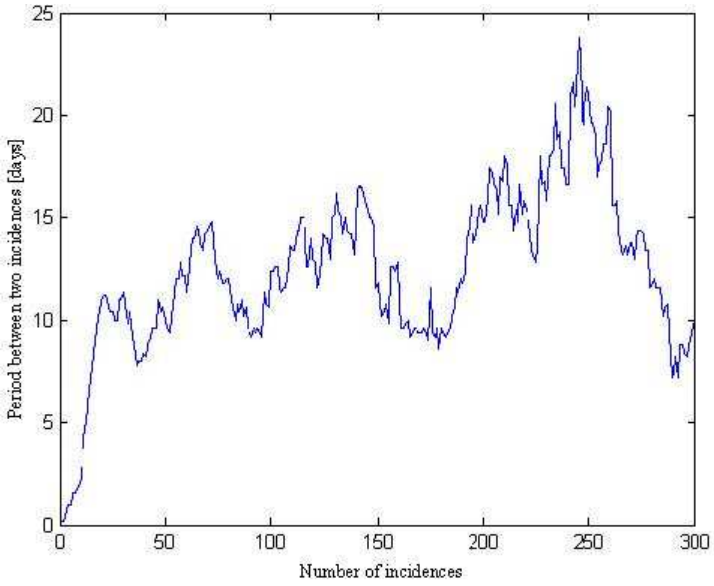


Figure 1: Driving data "in Oradea center"

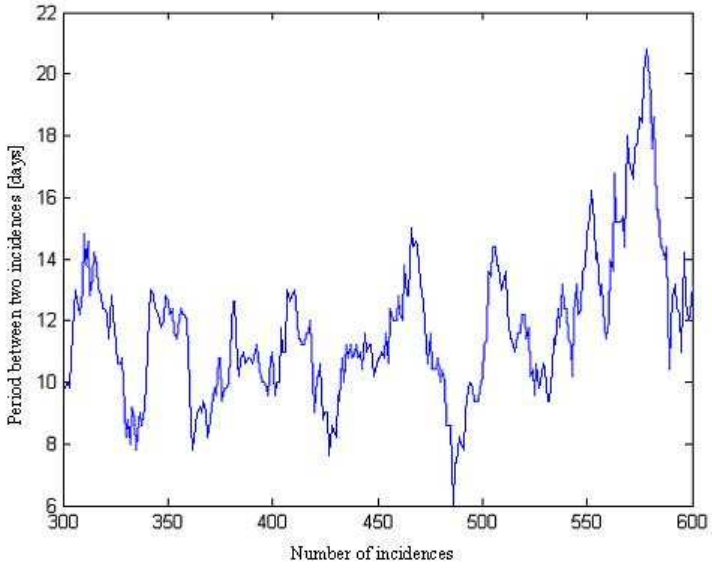


Figure 2: Checking data "in Oradea center"

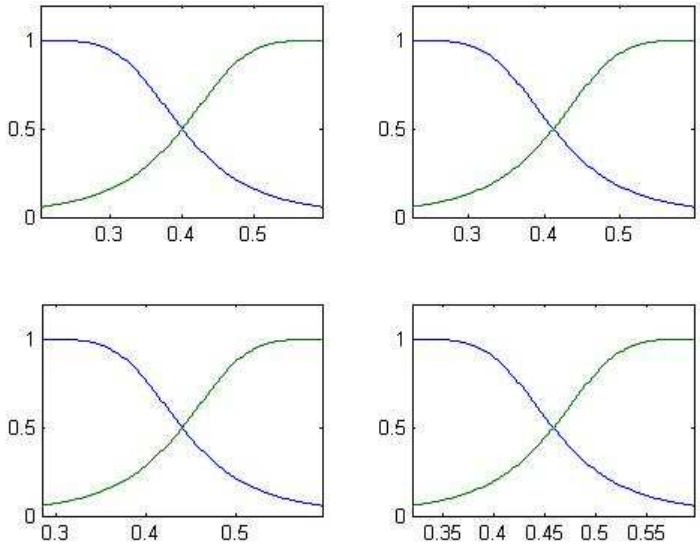


Figure 3: Initial membership function "in center Oradea"

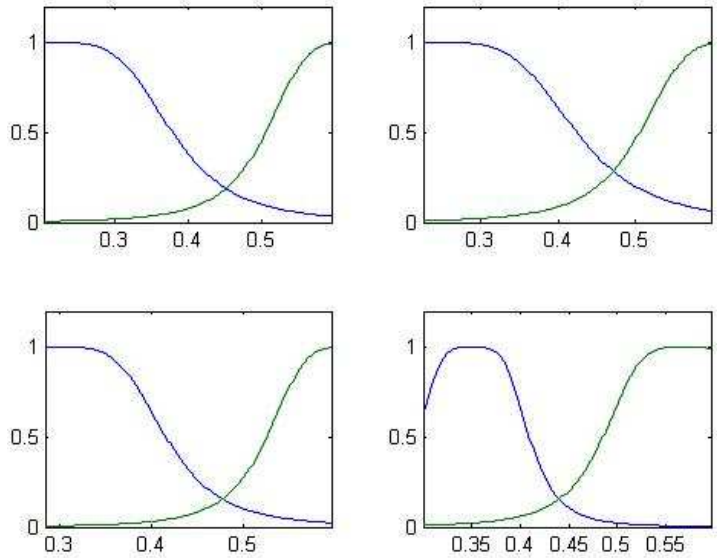


Figure 4: Adjusted membership functions "in center Oradea"

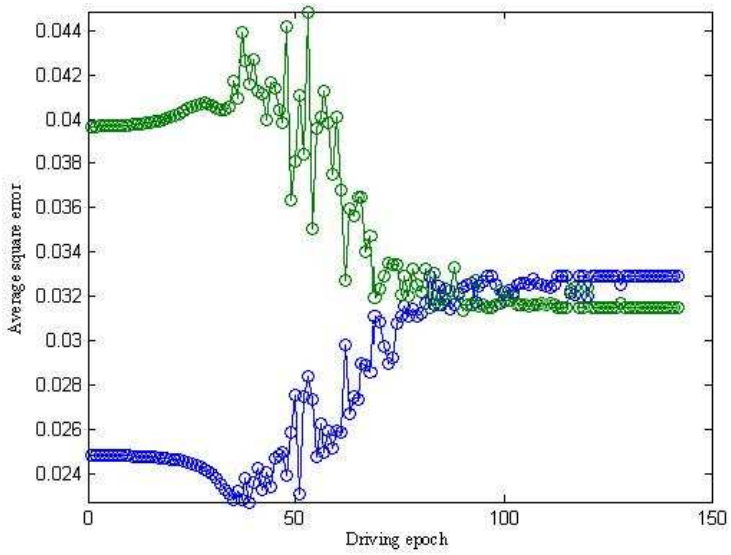


Figure 5: Comparing the driving errors (A) and checking it (V) "in center Oradea"

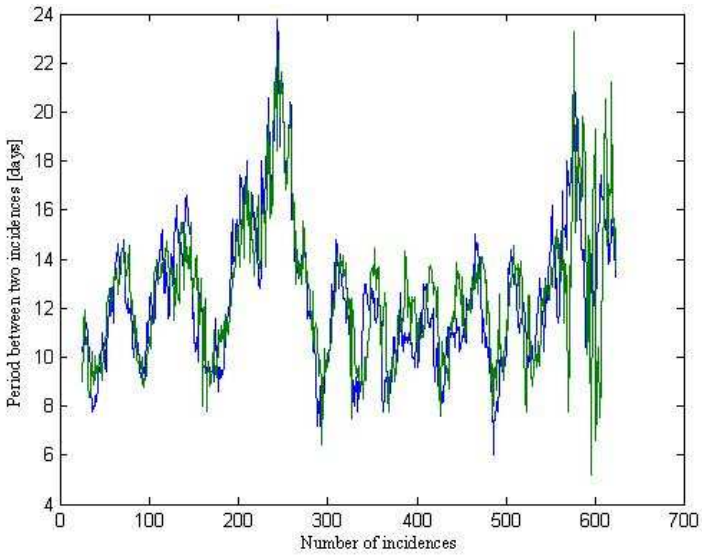


Figure 6: Evolution diagrams with real values (A), simulated values (V) "in center Oradea"

References

[1] Dzitac Simona, "Evaluation of quality indices of electric energy distribution service" - PhD Research No.3, December, University of Oradea, 2007. (in Romanian)

[2] Jang, J.-S. R., "Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm", *Proc. of the Ninth National Conf. on Artificial Intelligence (AAAI-91)*, pp. 762-767, July 1991.

[3] Jang, J.-S. R., "ANFIS: Adaptive-Network-based Fuzzy Inference Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp. 665-685, May 1993.

[4] Jang, J.-S. R. and C.-T. Sun, "Neuro-fuzzy modeling and control", *Proceedings of the IEEE*, March 1995.

[5] Jang, J.-S. R. and C.-T. Sun, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, 1997.

[6] Secui C., Felea I., Dzitac S., Dale E., Boja I., "Database and software system for evaluation of quality indexes of electric energy supplying service", Proceedings of the 7 th International Power Systems Conference, November 22-23, PSC 2007, pag. 589-596, ISSN 1582-7194. (in Romanian)

[7] Zadeh, L.A., "Fuzzy Logic", *Computer*, Vol. 1, No. 4, pp. 83-93, 1988.

[8] Research Report no. 3050/2006, University of Oradea, 2007.

Simona Dziţac, Ioan Felea, Tiberiu Vesselenyi
University of Oradea
Universitatii St. 1, 410087, Oradea, Romania
E-mail: sdzitac@rdslink.ro, ifelea@uoradea.ro, tvesselenyi@yahoo.co.uk

Ioan Dziţac
Agora University of Oradea
Department of Economic Informatics
Piata Tineretului 8, Oradea 410526, Romania
idzitac@univagora.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 293-298

Design and Implementation of DPA Resistive Grain-128 Stream Cipher Based on SABL Logic

R. Ebrahimi Atani, W. Meier, S. Mirzakuchaki, S. Ebrahimi Atani

Abstract: Cryptographic embedded systems are vulnerable to Differential Power Analysis (DPA) attacks. In this paper, Grain-128 stream cipher is implemented using SABL hiding technique. The cipher is designed by BSim 130nm technology and simulated by HSpice software. Simulations show DPA resistivity of SABL implementation of Grain-128 has a major improvement. The paper presents the tradeoffs involved in designing the architecture, the design for performance issues and the possibilities for future development.
Keywords: Stream cipher, Grain-128, SABL, Standard CMOS, DPA.

1 Introduction

Today, security in any form is an inevitable requirement for an increasing number of embedded systems, ranging from low end systems such as Personal Digital Assistants (PDA), wireless handsets, networked sensors, and smart cards to high-end network equipment such as routers, gateways, firewalls, storage and web servers. Technological advances that have spurred the development of these electronic systems have also ushered in seemingly parallel trends in the sophistication of security attacks. In 1998, Kocher et al. first reported that the power consumption of a smart card could reveal the secret key of the cryptographic algorithm [1]. The attack, called as Differential Power Analysis (DPA), has been considered as the most dangerous attack to the security of cryptographic embedded systems. DPA is a well-known and thoroughly studied threat for implementations of block ciphers (DES and AES), public key algorithms (RSA) and recently stream ciphers (Grain and Trivium). Stream ciphers as part of the symmetric key cryptography family, have always had the reputation of efficiency in hardware and speed. They have attracted much attention since the beginning of the eSTREAM project in 2004. Although there is vast literature about DPA on implementations of block ciphers and public key algorithms, only few publications can be found about attacks on stream ciphers. In [2] side-channel analysis and their applicability on stream ciphers has been surveyed. In [3], a theoretical known IV DPA attack on $A5/1$ (used in GSM) and E_0 (used in Bluetooth) has been described. In power analysis attacks, it is assumed that the power consumption of a circuit is correlated to the data handled. An attacker can therefore recover secret information by simply monitoring the power signals of a running device. Stream ciphers require frequent synchronization to prevent synchronization loss between sender and receiver. Normally the initialization will be done with the same secret key and with a different initial value IV. So an attacker can disrupt the synchronization and apply a new known IV and measure the power traces in initialization phase to apply a DPA on the embedded system of the stream cipher. So far, there is only one report on a practical DPA targeting hardware implementations of stream ciphers [4]. In that paper a chosen IV DPA attack on Grain and Trivium stream ciphers has been described and executed. The attack is based on a novel concept of choosing initial value vectors, so that the algorithmic noise of the device is totally eliminated. Protecting implementations against DPA attacks is usually difficult and expensive. The goal of countermeasures against DPA attacks is to make the power consumption independent of intermediate values of the stream cipher. There are three basic groups into which these countermeasures can be characterized: protocols, masking and hiding [7]. In this paper hiding techniques is used to remove the data dependency of power consumption. This means the characteristics of power consumption of the device are changed so that an attacker cannot easily find a data dependency. It is important to point out that implementations protected by hiding countermeasures process the same intermediate results as unprotected implementations. The principles of the countermeasures can be implemented at different levels in a cryptographic device. In general, these techniques are theoretical countermeasures and only reduce the side channel leakage and do not fundamentally prevent a DPA (small power variations still appear in function of the input sequences). But the advantage of these countermeasures is to make the attack significantly harder.

In this article, we provide a brief overview of hiding countermeasures. So far a number of logics have been proposed for the realization of hiding countermeasures. However, there has not been a unified architecture which can be used as a test bench for applicability of these logic styles on stream ciphers. In this paper, we will exploit Sense Amplifier Based Logic (SABL) to counteract power analysis in Grain stream cipher submitted to eSTREAM in 2005. Power traces of the resulting circuits exhibit that SABL significantly reduces the signal to noise ratio (SNR).

This paper is structured as follows. A general model of power analysis attack on stream ciphers is given in Section 2. Section 3 describes an overview of hiding countermeasures. In section 4 The Grain-128 will be explained. Experiments and results based on the design of Grain are in Section 5 and finally, conclusions are in Section 6.

2 Differential Power Analysis of Stream Ciphers

Differential Power Analysis is the most powerful side channel attack. The attack is based on the fact that CMOS logic and application specific details cause logic operations to have power characteristics that depend on the input data. It relies further on statistical analysis and error correction to extract the information from the power consumption that is correlated to the secret key [1].

In a DPA a hypothetical model of the device under attack is used to predict the power consumption. Next a set of hypotheses about the intermediate values is generated. An attacker tries to identify the true hypothesis by finding the highest correlation between the power consumption of the physical realization of an algorithm and those internal bits which can be computed by the attacker by virtue of one of these hypotheses. The quality of the model has a strong impact on the effectiveness of the attack and it is therefore of primary importance. This model is typically not very complex. The classical setup for a DPA on stream ciphers is illustrated in Fig. 1. On the left side we have the cryptographic embedded system of a stream cipher to be attacked. Output power traces are determined by the input data, IV, private key, output of the device and by many other parameters. An attacker to some extent has the potential knowledge of some of them (e.g. IV, input data and output data) while others are not. Regarding a DPA attack, multiple measurements of the power consumption of a cryptographic device are made. For each measurement, different chosen IV's are sent to the device. Since the cryptographic algorithm is known, a key hypothesis can be used to calculate the targeted data values based on the random input values. If the correct key hypothesis is used, the targeted data values are calculated correctly for all measurements. According to (1) total power consumption of an embedded device depends on 3 factors:

$$P_{Total} = P_{Cons.} + P_{Noise} + P_{DD} \quad (1)$$

With the help of statistical methods (calculation of correlations, mean values, etc.), the randomness of the data values that are not targeted($P_{Const.}$: leakage currents and data independent power consumption and P_{Noise}: which comes from electrical noise) is exploited to reduce their effects on the power consumption traces. P_{DD} is the data dependent power consumption and is targeted in statistical analysis.

After all, the result of the statistical operation indicates which key hypothesis is correct. The model of the side channel used by the attacker is shown in Fig.1. The model may consider additional parameters besides the key, the input and the output of the module. There are two generic power models that are commonly used for power analysis attacks: Hamming weight and Hamming distance. Hamming distance power model which can describe DPA better, is used to map the transitions that occur at the outputs of cells of a net list to power consumption values. For example, in CMOS gates, it is reasonable to assume that the main component of the data dependent power consumption is the dynamic power consumption which is the power dissipation of charging and discharging of output capacitance nodes($P_{0\rightarrow1}$ or $P_{1\rightarrow0}$). In a CMOS gate, we can express dynamic power consumption by:

$$P_{Dynamic} = N_{Switching} \cdot C_L \cdot f \cdot V_{Supply}^2 \quad (2)$$

Where C_L is the gate load capacitance and $N_{Switching}$ is the probability of a $0 \rightarrow 1$ or $1 \rightarrow 0$ output transition and f is the clock frequency. This equation shows that the power consumption of CMOS circuits is data dependent. An attacker may consequently estimate power consumption of a device at time t by the number of bit transitions inside the device at this time and use it as hypothetical model for generation of required hypotheses sets. Note that $N_{Switching}$ is the most important factor in the hypothetical model. There are different techniques for calculation of it. For example variable gate delay model can be used for measuring the number of transitions and glitches of a circuit. this technique can be easily applied to circuits using a VHDL simulator in Register Transfer level.

3 Hiding Logic Styles

The first structured approach to counteract DPA attacks at the cell level was the use of hiding logic styles. These styles try to break the correlation between an algorithm's intermediate results and the power consumption of

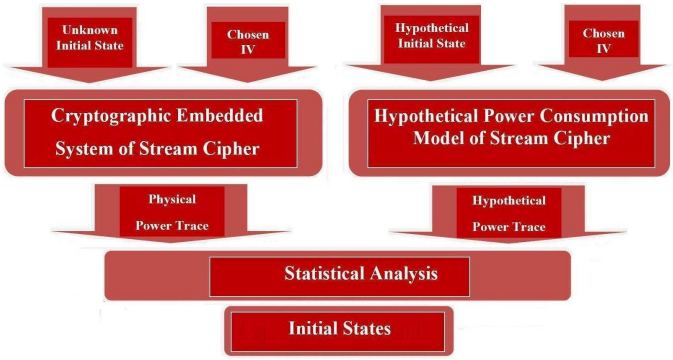


Figure 1: Differential power analysis model of stream ciphers.

the cryptographic device that executes this algorithm by making the instantaneous power consumption of the cells either random or the same in each clock cycle.

The three major types of hiding logic styles are: Dual rail Precharge (DRP), Asynchronous, and Current Mode Logic (CML). DRP logic styles are the most popular type. As the name implies, the concepts of dual rail and precharge logic are combined to achieve constant power consumption. Precharging breaks a signal's sequence of values by splitting each clock cycle into precharge and evaluation phases. In the precharge phase, the complementary wires encoding a signal are set to a predefined precharge value, such as 1. In the subsequent evaluation phase, one of the two complementary wires is set to 1 according to the actual value that is processed. As a result, for each signal in a circuit, exactly one $0 \rightarrow 1$ transition and one $1 \rightarrow 0$ transition occur in a clock cycle. By ensuring a balance between the complementary wires between cells on the one hand and a balance of the internal structure of the cells on the other hand, designers can achieve constant power consumption. Examples of DRP logic styles are Sense Amplifier Based Logic (SABL), Wave Dynamic Differential Logic (WDDL), Dual Spacer Dual Rail Logic (DSDR), Three Phase Dual Rail Precharge Logic (TDPL), and Three State Dynamic Logic (3sDL). In this paper we will concentrate on SABL [6] for implementation of Grain. Fig.2 shows the transistor schematic of SABL gates. As can be seen in the Fig.2 SABL gates can be designed using DPDN or DPUN, respectively controlled by clk and $\overline{clk}$. This allows two modes for cascading SABL gates: domino connection (by connecting the outputs of the gate to the inputs of the next gate through inverters) or np-connection (n-gates followed by p-gates like in NP-logic). The second type of hiding logic style is asynchronous logic. Asynchronous randomization is inherently data dependent and thus does not lead to high security. High complexity of design flow in asynchronous logic is another disadvantage of using them. As a result, asynchronous logic styles typically resort to the DRP principle to achieve constant power consumption. The last type of hiding logic style is CML, in which logic values are encoded by current flows that take different paths in a circuit. CML gates have the advantage of high speed and low noise but also suffer from a complex design flow and static power dissipation. As an example, Dynamic Current Mode Logic (DyCML) is shown in Fig.2 (iii). In order to evaluate the effect of SABL onto the

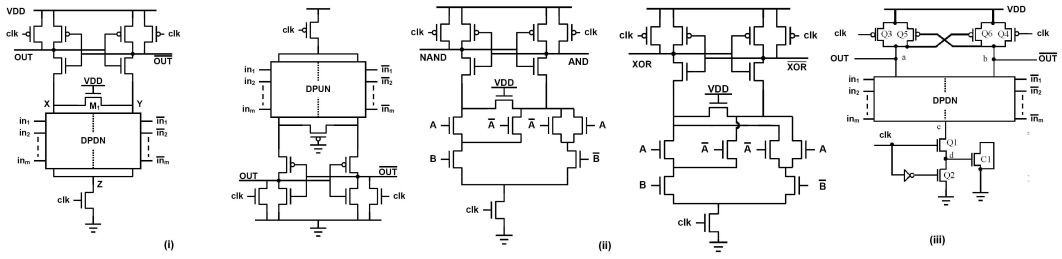


Figure 2: (i) SABL logic styles (ii) SABL gates Nand2(left) Xor2(right) (iii) DyCML logic styles .

resistance against power analysis, we compared it with standard CMOS in implementation of a 4-input Xor using three Xor2. Fig. 3 shows the resulting Xor4 output and its supply current in both circuits. In SABL part, power consumption in each cycle is equal. As can be seen standard CMOS cannot filter glitches and there is a peak in power consumption in 4 μs. One of the drawbacks of using SABL is the supply current spikes which appear at the beginning of the precharge phase. During the design of Grain we will present a technique to minimize these

spikes.

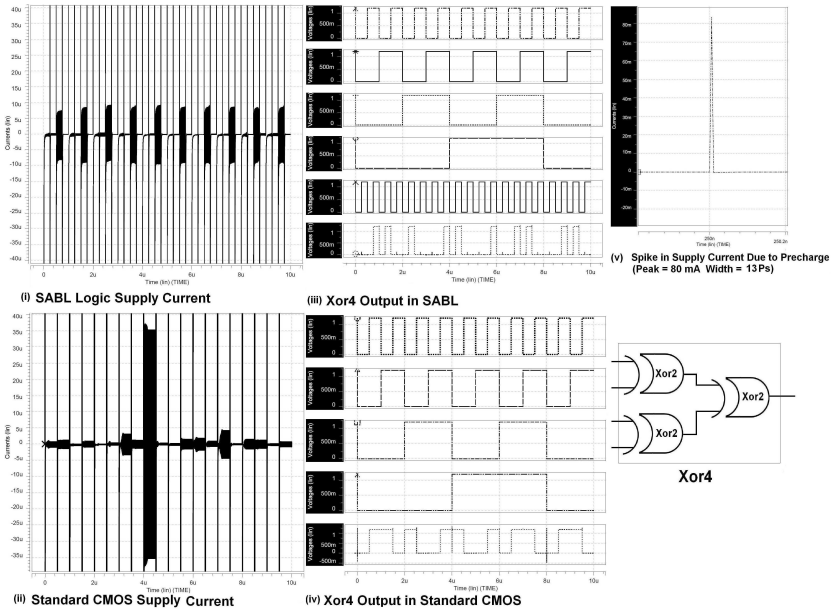


Figure 3: (i) SABL supply current (ii) CMOS supply current (iii) SABL Xor4 output (iv) CMOS Xor4 output (v) SABL current spike.

4 Grain Stream Cipher

Grain 128 [5] is a stream cipher introduced in 2005 as a candidate for the hardware profile of eSTREAM project, that has been selected in April 2007 for the third and ultimate phase of the competition. Grain-128 is a binary additive synchronous stream cipher with an internal state of 256 bits s_i, $s_{i+1},\ldots$, s_{i+127} and b_i, $b_{i+1},\ldots$, b_{i+127} residing in a linear feedback shift register (LFSR) and a nonlinear feedback shift register (NLFSR), respectively. The design of the algorithm mainly targets hardware environments where gate count, power consumption and memory is very limited. In general, a stream cipher consists of two phases. The first phase is the initialization of the internal state using the secret key and the IV. After that, the state is repeatedly updated and hence used to generate key-stream bits. The key size of Grain is 128 bits (k_i, $0 \leq i \leq 127$). Additionally, an initial value of 96 bits (IV_i, $0 \leq i \leq 95$) is required. The initialization of the key and IV is done as follows. The 128 NLFSR elements are loaded with the key bits, ($b_i = k_i$, $0 \leq i \leq 127$), then the first 96 LFSR elements are loaded with the IV bits, ($s_i = IV_i$, $0 \leq i \leq 95$). The last 32 bits of the LFSR are filled with ones, $s_i = 1$, $96 \leq i \leq 127$. The basic structure of the algorithm and key initialization of Grain can be seen in Fig. 3. Two polynomials of degree 128, $f(x)$ and $g(x)$, are used as feedback function for the LFSR and NLFSR.

$$f : s_{i+128} = s_i \oplus s_{i+7} \oplus s_{i+38} \oplus s_{i+70} \oplus s_{i+81} \oplus s_{i+96}$$
$$g : b_{i+128} = s_i \oplus b_i \oplus b_{i+26} \oplus b_{i+56} \oplus b_{i+91} \oplus b_{i+96} \oplus b_{i+3} \cdot b_{i+67} \oplus b_{i+11} \cdot b_{i+13}$$
$$\oplus b_{i+17} \cdot b_{i+18} \oplus b_{i+27} \cdot b_{i+59} \oplus b_{i+40} \cdot b_{i+48} \oplus b_{i+61} \cdot b_{i+65} \oplus b_{i+68} \cdot b_{i+84}$$

The output function $h(x)$ uses as input selected bits from both feedback shift registers:

$$h(x) = b_{i+12} \cdot s_{i+8} \oplus s_{i+13} \cdot s_{i+20} \oplus b_{i+95} \cdot s_{i+42} \oplus s_{i+60} \cdot s_{i+79} \oplus b_{i+12} \cdot b_{i+95} \cdot s_{i+95}$$

Additionally, seven bits of NLFSR are XORed together and the result is added to the function $h(x)$. This output is used during the initialization phase as additional feedback to LFSR and NLFSR. During normal operation this value is used as key stream output.

$$z_i = s_{i+93} \oplus h(x) \oplus [b_{i+2} \oplus b_{i+15} \oplus b_{i+36} \oplus b_{i+45} \oplus b_{i+64} \oplus b_{i+73} \oplus b_{i+89}]$$

5 Design and Simulation Results of Grain-128

The whole cipher including the control unit for initialization phase is modeled at transistor level using spice net list and typical BSIM3 0.13 μm technology for $V_{dd} = 1.2V$. This technology has a threshold voltage of respectively 0.326V and -0.324V for nMOS and pMOS devices. Spice simulations were run to test the circuit by test vectors provided by inventors of Grain using Hspice circuit simulator and C compiler. Domino cascading scheme has been used for all SABL gate connections. 256 internal states of Grain (LFSR and NLFSR) have been designed by SABL flip-flops. In the initialization of Grain, private key and initial values can be loaded in parallel or sequentially. In parallel loading key and IV will be loaded in the LFSR and NLFSR after the first rising edge of the *Clk* signal, and the whole initialization needs 257 cycles with the cost of 3 more Nand2 for each FlipFlop. In case of sequential loading after 128 *Clk* pulses for inputting the key and IV, another 256 *clk* pulse is needed and initialization takes more time but less Silicon. The loading of data is done by resetting $\overline{Load}/shift$ line of architecture. Since all the components in the architecture need a *Clk* signal for switching from Precharge phase into evaluation and vice versa, a chained buffer *clk* signal is needed. Design is done by minimum size transistors in the technology in order to lower the total capacitance to get lower Dynamic power in (2). Besides, this will help to cut the current spikes in the beginning of the precharge phase of each cycle. These spikes which are one of the disadvantages of using SABL can damage the voltage source (e.g. battery) of the chip. The width of this spikes is a factor of maximum current output of the voltage source and normally it is less than 1ns Fig. 4 (iii-a). In order to get rid of the spikes we have used delayed clock mechanism. So NLFSR will start to precharge after LFSR. This technique will not disrupt the output of the cipher. In case of 1ns delay the peak of current spike is half which is shown in Fig. 4 (iii-b). Supply current traces of the architecture for SABL design and standard CMOS design in initialization phase is shown in Fig. 4. This simulation is for $key = AAA\ldots AAA(128bits)$ and two initial values $IV_1 = FFF\ldots FFF, IV_2 = 888\ldots 888(96bits)$. As can be easily seen in the figure the power consumption of SABL design for different IV's is the same and to some extent we can say in SABL in all cycles the power consumption is constant. For precharging of each Cycle a total Charge of 600 pQ is needed. All power simulations is observed by 3.3 *Mhz* Clock.

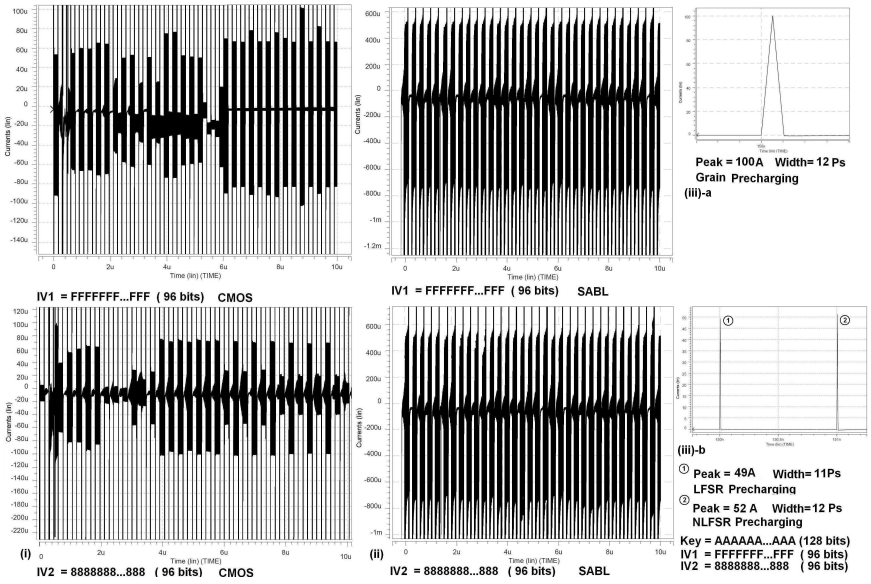


Figure 4: (i) Standard CMOS design of Grain (ii) SABL design of Grain (iii) Spikes in precharge

6 Summary and Conclusions

This paper investigated the use of SABL logic to counteract power analysis attacks. In particular, an efficient DPA resistive circuit for Grain-128 stream cipher has been designed and compared with the standard CMOS implementation of Grain. First we exhibited that SABL allow to significantly decrease the circuit energy variations. This is due to equal amounts of power consumption in each clock cycle of SABL gates. This implementation has

been done on transistor level but in practice the cipher itself is part of a system on chip with lots of other circuits which can increase $P_{Cons.} + P_{Noise}$ in (1) to achieve lower SNR. Although SABL cannot be completely tamper resistant, this logic probably presents acceptable security margins for general applications of Grain. For future work interested researchers can investigate some circuit changes in SABL styles to counteract other side channel attacks such as fault attacks to obtain more security.

Acknowledgment

Reza Ebrahimi Atani and Sattar Mirzakuchaki wish to thank the Iran Telecommunication Research Center (ITRC) for their financial support.

References

[1] P. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," *Advances in Cryptology - CRYPTO'99*, Springer-Verlag, LNCS Vol. 1666, pp. 388–397, 1999.

[2] Ch. Rechberger and E. Oswald, "Stream Ciphers and Side-Channel Analysis" *In SASC 2004 - The State of the Art of Stream Ciphers*, Brugge, Belgium, Workshop Record, pp. 320–326, Oct. 14-15, 2004.

[3] J. Lano, N. Mentens, B Preneel, and I. Verbauwhede, "Power Analysis of Synchronous Stream Ciphers with Resynchronization Mechanism" *In SASC 2004 - The State of the Art of Stream Ciphers*, Brugge, Belgium, Workshop Record, pp. 327–333, Oct. 14-15, 2004.

[4] W. Fischer, B. M. Gammel, O. Kniffler, J. Velton, "Differential Power Analysis of Stream Ciphers," *Topics in Cryptology - CT-RSA 2007*, Springer-Verlag, LNCS, Vol. 4377, pp. 257–270, 2007.

[5] M. Hell, Th. Johansson, A. Maximov, and W. Meier, "Grain - A Stream Cipher for Constrained Environments," 2006. Available at http://www.ecrypt.eu.org/stream/p3ciphers/grain/Grain128 p3.pdf.

[6] K. Tiri, M. Akmal, and I. Verbauwhede, "A Dynamic and Differential CMOS Logic with Signal Independent Power Consumption to Withstand Differential Power Analysis on Smart Cards," *28th European Solid State Circuits Conference*, IEEE Press, pp. 403 – 406, , 24-26 Sep. 2002.

[7] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards,* Springer, 2007.

R.Ebrahimi Atani, W. Meier, S. Mirzakuchaki, S.Ebrahimi Atani
IUST, FHNW, IUST, Guilan University
Electrical Engineering, IAST, Electrical Engineering, Mathematics Department
E-mail: rebrahimi@iust.ac.ir, willi.meier@fhnw.ch, m-kuchaki@iust.ac.ir, ebrahimi@guilan.ac.ir

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 299-303

Approximating the Periodic Solution of Delay Volterra Integral Equation from Biomathematics

Loredana-Florentina Galea

Abstract: In periodic and Lipschitz conditions, a numerical method to approximate the periodic solution of delay Volterra nonlinear integral equation arising in biomathematics is obtained. The method combines the sequence of successive approximations with recent trapezoidal quadrature rule for Lipschitzian functions obtaining effective apriori error estimation.
Keywords: delay Volterra integral equations, periodic solutions, successive approximations, trapezoidal quadrature rule.
2000 AMS Mathematics Subject Classification: 45D05, 65D32, 92C60.

1 Introduction

As in [5], [1] and [6], it's consider an isolated population. We suppose that:

(i) the population have a constant number of persons, namely is in the neighborhood of population by stable poise type;

(ii) the population is composed by two different classes: susceptible persons of infection and infected persons;

(iii) infection not lead to demise and not offer immunity;

(iv) the length of time in which a person remains infections is constant, and is denoted by τ.

Knowing the proportion of infections in population at initial time t_0, it's demand to determinate the proportion of infections in population at time t. Here, $x(t)$ is the proportion of infections in population at time t, and $f(t,x(t))$ is the proportion of new infectivities at time t. In these conditions, the mathematical model of the problem is the following delay integral equation:

$$(1)\ x(t) = \int_{t-\tau}^{t} f(s,x(s))\,ds$$

This equation is a model for the spread of certain infectious disease with contact rate that varies seasonally. Also, the model is applied in the study of the growth of population in a certain environment with characteristics that vary periodical in time. A presentation of this model can be found in [3]. This model can be used also for the study of the growth of population in an environment that vary periodical, case when from the equation (1) is a vector valued function.

Thus, in what follow, we consider the Volterra nonlinear delay integral equations:

$$x(t) = \int_{t-\tau}^{t} f(s,x(s))\,ds$$

We suppose that $f \in C(\mathbb{R},\mathbb{R}_+)$ and exists $\omega > 0$ such that:

$$f(t+\omega,x) = f(t,x),\ \forall\, t \in \mathbb{R}$$

We consider the following functional spaces:

$$X(\omega) = \{x : \mathbb{R} \to \mathbb{R} | x \in C(\mathbb{R}),\ x(t+\omega) = x(t),\ \forall\, t \in \mathbb{R}\}$$

and $X_+(\omega) = \{x \in X(\omega) | x(t) \geqslant 0,\ \forall\, t \in \mathbb{R}\}$ which is closed set in $X(\omega)$ and $X(\omega)$ is a generalized metric space with:

$$d_C : X \times X \to \mathbb{R}^2,\ d_C((x_1,y_1),(x_2,y_2)) = (\|x_1 - x_2\|, \|y_1 - y_2\|)$$

where $\|u\| = \max\{|u(t)| : t \in [0,\omega]\}$, for any $u \in X(\omega)$.

To obtain the existence and uniqueness of the periodic solution of the equation (1), we use the Banach's principle and Picard's theorem.

2 Existence and uniqueness of the solution

Consider the following function $A : X_+(\omega) \to C(\mathbb{R})$ defined by:

$$A(x)(t) = \int_{t-\tau}^{t} f(s, x(s))\, ds$$

We will consider the conditions:
(i) (continuity): $f \in C(\mathbb{R} \times \mathbb{R}_+)$;
(ii) (boundedness): exists $m, M \geqslant 0$ such that:

$$0 \leqslant m \leqslant f(t,u) \leqslant M,\ \forall\ (t,u) \in \mathbb{R} \times \mathbb{R}_+;$$

(iii) (periodicity): exists $\omega > 0$ such that:

$$f(t+\omega, u) = f(t,u),\ \forall\ (t,u) \in \mathbb{R} \times \mathbb{R}_+$$

(iv) (first Lipschitz condition): exist $L > 0$ such that:

$$|f(t,u) - f(t,v)| \leqslant L|u-v|,\ \forall\, t \in \mathbb{R},\ \forall\, u, v \in \mathbb{R}_+$$

(v) (second Lipschitz condition): exist $\gamma > 0$ such that:

$$|f(t_1,u) - f(t_2,u)| \leqslant \gamma |t_1 - t_2|,\ \forall\, t_1, t_2 \in \mathbb{R},\ \forall\, u \in \mathbb{R}_+$$

(vi) $L\tau < 1$

Theorem 1. *If the conditions (i)-(iv) and (vi) are satisfied then the equations (1) has an unique solution in* $X_+(\omega)$.

Proof. We prove that $A(X(\omega)) \subset X(\omega)$. So, we have:

$$A(x)(t+\omega) = \int_{t+\omega-\tau}^{t+\omega} f(s, x(s))\, ds = \int_{t-\tau}^{t} f(u-\omega, x(u-\omega))du =$$

$$= \int_{t-\tau}^{t} f(u-\omega+\omega, x(u-\omega+\omega))du = A(x)(t),\ \forall\, t \in \mathbb{R},\ \forall\, x \in X(\omega)$$

Thus, $A(X(\omega)) \subset X(\omega)$ and $f(x(t)) \geqslant 0,\ \forall\, t \in \mathbb{R},\ \forall\, x \in X_+(\omega)$.
Let $x, y \in X(\omega)$ and we have:

$$|A(x)(t) - A(y)(t)| = \left| \int_{t-\tau}^{t} f(s, x(s))\, ds - \int_{t-\tau}^{t} f(s, y(s))\, ds \right| \leqslant$$

$$\leqslant L\tau \|x - y\|, \forall t \in \mathbb{R}$$

Therefore, $d_C(A(x), A(y)) \leqslant L\tau d_C(x,y)$.
Because $L\tau < 1$, we have that A has a fixed point on $X_+(\omega)$, denote by x^*. From condition (i), follow that $x^* \in C^1(\mathbb{R})$. □

Remark 2. Suppose that there exists $\varphi \in C(\mathbb{R})$ such that $x(0) = \varphi(0)$. It is easy to see that, in the conditions of Theorem 1, the terms of sequence of successive approximations given by:
$x_0(t) = \varphi(0),\ \forall\, t \in \mathbb{R}$
$x_m(t) = \int_{t-\tau}^{t} f(s, x_{m-1}(s))\, ds,\ t \in \mathbb{R},\ m \in \mathbb{N}^*$
are periodic functions. Indeed, x_0 is periodic (trivially constant),
$x_1(t+\omega) = \int_{t+\omega-\tau}^{t+\omega} f(s, \varphi(0))\, ds = \int_{t-\tau}^{t} f(u+\omega, \varphi(0))\, du =$
$= \int_{t-\tau}^{t} f(u, \varphi(0))du = x_1(t),\ \forall\, t \in \mathbb{R}$
and by induction, in the same way, follows that:
$x_m(t+\omega) = x_m(t),\ \forall\, t \in \mathbb{R},\ \forall\, m \geqslant 2.$

3 Approximation of the solution

We suppose that exists $l \in \mathbb{N}^*$ such that $\omega = l\tau$.
Let

$$\Delta : 0 = t_0 < t_1 < ... < t_n = \tau < ... < t_{q-1} < t_q = \omega$$

an equidistant division of $[0, \omega]$, where:

$$t_{-i} = -i\tfrac{\tau}{n},\ \forall\, i = \overline{0,n} \text{ and } t_i = i\tfrac{\tau}{n},\ \forall\, i = \overline{n+1,q}$$

with $q = nl$.

For equation $x(t) = \int\limits_{t-\tau}^{t} f(s, x(s))\, ds,\ \forall\, t \in \mathbb{R}$ the sequence of successive approximation is given by:

$$(2)\ \begin{cases} x_0(t) = \varphi(t) = \varphi(0),\ \forall\, t \in \mathbb{R} \\ x_m(t) = \int\limits_{t-\tau}^{t} f(s, x_{m-1}(s))\, ds,\ m \in \mathbb{N},\ \forall\, t \in \mathbb{R} \end{cases}$$

$x_m(t) = A(x_{m-1}),\ \forall\, m \in \mathbb{N}$, which on the knots of division Δ, become:

$$(3)\ x_m(t_i) = \int\limits_{t_i-\tau}^{t_i} f(s, x_{m-1}(s))\, ds,\ \forall\, i = \overline{0,q}$$

Consider the function $F_m : [-\tau, \omega] \to \mathbb{R},\ F_m(t) = f(t, x_m(t)),\ \forall\, m \in \mathbb{N}$.

Then, $x_m(t) = \int\limits_{t-\tau}^{t} F_{m-1}(s) ds,\ \forall\, t \in \mathbb{R},\ \forall\, m \in \mathbb{N}$ which on the knots of division Δ, becomes:

$$x_m(t_i) = \int\limits_{t_i-\tau}^{t_i} F_{m-1}(s) ds,\ \forall\, i = \overline{0,q},\ \forall\, m \in \mathbb{N}.$$

For the calculus of the integrals (2) we apply the trapezoid quadrature rule for Lipschitzian functions

$$(4)\ \int\limits_a^b f(t)\, dt = \tfrac{b-a}{2n}\left[f(a) + 2\sum_{i=1}^{n-1} f(t_i) + f(b)\right] + R_n(f)$$

where the remainder $R_n(f)$ satisfies the inequality:

$$|R_n(f)| \leqslant \tfrac{(b-a)^2}{4n} \cdot L.$$

Using the quadrature rule (3), we have:

$$(5)\ \begin{cases} x_m(t_i) = \dfrac{\tau}{2n}\left[f(t_i - \tau, x_{m-1}(t_{q+i} - \tau)) + 2\sum\limits_{j=1}^{n-1} f(t_{i+j} - \tau, x_{m-1}(t_{q+j} - \tau)) + \right. \\ \left. + f(t_i, x_{m-1}(t_i))\right] + R_{m,i},\ \forall\, i = \overline{0,n} \\ x_m(t_i) = \dfrac{\tau}{2n}\left[f(t_i - \tau, x_{m-1}(t_i - \tau)) + 2\sum\limits_{j=1}^{n-1} f(t_{i+j} - \tau, x_{m-1}(t_{i+j} - \tau)) + \right. \\ \left. + f(t_i, x_{m-1}(t_i))\right] + R_{m,i},\ \forall\, i = \overline{n+1,q} \end{cases}$$

that is,

$$(6)\ \begin{cases} x_m(t_i) = \dfrac{\tau}{2n}\left[f(t_{i-n}, x_{m-1}(t_{q-n+i})) + 2\sum\limits_{j=1}^{n-1} f(t_{i-n+j}, x_{m-1}(t_{q-n+j})) + \right. \\ \left. + f(t_i, x_{m-1}(t_i))\right] + R_{m,i},\ \forall\, i = \overline{0,n},\ \forall\, m \geqslant 2 \\ x_m(t_i) = \dfrac{\tau}{2n}\left[f(t_{i-n}, x_{m-1}(t_{i-n})) + 2\sum\limits_{j=1}^{n-1} f(t_{i-n+j}, x_{m-1}(t_{i-n+j})) + \right. \\ \left. + f(t_i, x_{m-1}(t_i))\right] + R_{m,i},\ \forall\, i = \overline{n+1,q},\ \forall\, m \geqslant 2 \end{cases}$$

Relations (5) and (6) lead us to the following algorithm:

$$x_0(t_i)=\varphi(0),\ \forall\, i=\overline{0,q}$$
$$x_1(t_i)=\frac{\tau}{2n}\left[f(t_i-\tau,\varphi(0))+2\sum_{j=1}^{n-1}f(t_{i+j}-\tau,\varphi(0))+f(t_i,\varphi(0))\right]+R_{1,i}(f)+$$
$$=\frac{\tau}{2n}\left[f(t_{i-n},\varphi(0))+2\sum_{j=1}^{n-1}f(t_{i-n+j},\varphi(0))+f(t_i,\varphi(0))\right]+R_{1,i}(f)$$

By induction, for $m\geqslant 2$, it follows:

$$x_m(t_i)=\frac{\tau}{2n}\left[f\left(t_{i-n},\overline{x_{m-1}(t_{q-n-i})}\right)+2\sum_{j=1}^{n-1}f\left(t_{i-n+j},\overline{x_{m-1}(t_{q-n-j})}\right)+\right.$$
$$\left.+f\left(t_{i-n+j},\overline{x_{m-1}(t_i)}\right)\right]+\overline{R_{m,i}(f)}=\overline{x_m(t_i)}+\overline{R_{m,i}(f)},\ \forall\, i=\overline{0,n}$$
$$x_m(t_i)=\frac{\tau}{2n}\left[f\left(t_{i-n},\overline{x_{m-1}(t_{i-n})}\right)+2\sum_{j=1}^{n-1}f\left(t_{i-n+j},\overline{x_{m-1}(t_{i-n-j})}\right)+\right.$$
$$\left.+f\left(t_i,\overline{x_{m-1}(t_i)}\right)\right]+\overline{R_{m,i}(f)}=\overline{x_m(t_i)}+\overline{R_{m,i}(f)},\ \forall\, i=\overline{n+1,q}$$

In what follows, we have:

$$|F_0(t_1)-F_0(t_2)|\leqslant\gamma|t_1-t_2|$$
$$|F_m(t_1)-F_m(t_2)|\leqslant|t_1-t_2|(\gamma+2ML),\ \forall\, t_1,t_2\in[0,\omega],\ m\in\mathbb{N}^*$$

and if $M\geqslant 0$ then $|f(t,u)|\leqslant M,\ \forall\, t\in[0,\omega],\ \forall\, u\in\mathbb{R}$.
In this way we have obtain:

Theorem 3. *Consider the functions*
$F_m:[-\tau,\omega]\to\mathbb{R},\ F_m(t)=f(t,x_m(t)),\ \forall\, t\in[-\tau,\omega],\forall\, m\in\mathbb{N}$.
Suppose that:
(i) f - continuous;
(ii) exists $L,\gamma>0$ such that:
$|f(t,u_1)-f(t,u_2)|\leqslant L|u_1-u_2|,\ \forall\, u_1,u_2\in\mathbb{R}$ *and*
$|f(t_1,u)-f(t_2,u)|\leqslant\gamma|t_1-t_2|,\ \forall\, t_1,t_2\in[0,\omega],\ \forall\, u\in\mathbb{R}$;
(iii) exist $M\geqslant 0$ such that $|f(t,u)|\leqslant M,\ \forall\, t\in[0,\omega],\ u\in\mathbb{R}$.
Then all functions $F_m,\ m\in\mathbb{N}$ are Lipschitzian with the Lipschitz constant $\gamma+2ML$.

4 The convergence of the method and main result

From Theorem 2, the error estimation of the remainders is

$$|R_{m,i}(f)|\leqslant\frac{\tau^2}{4n}(\gamma+2ML),\ \forall\, m\in\mathbb{N}^*,\ i=\overline{1,n}$$

where f is Lipschitz in respect to both of arguments.
From relation (7) we obtain the estimations:

$$\left|\overline{R_{2,i}(f)}\right|=\left|x_2(t_i)-\overline{x_2(t_i)}\right|\leqslant$$
$$\leqslant\frac{\tau^2}{4n}(\gamma+2ML)+\tau L\frac{\tau^2(\gamma+2ML)}{4n}=\frac{\tau^2}{4n}(\gamma+2ML)(1+L\tau),\ i=\overline{0,n}$$

From relation (8) we obtain the estimations:

$$\left|\overline{R_{2,i}(f)}\right|=\left|x_2(t_i)-\overline{x_2(t_i)}\right|\leqslant$$
$$\leqslant\frac{\tau^2}{4n}(\gamma+2ML)+\tau L\frac{\tau^2(\gamma+2ML)}{4n}=\frac{\tau^2}{4n}(\gamma+2ML)(1+L\tau),\ i=\overline{n+1,q}$$

By induction, for $m\geqslant 3$, we obtain:

$$\left|\overline{R_{m,i}(f)}\right| \leqslant \frac{\tau^2}{4n}(\gamma+2ML)\left(1+L\tau+...+L^{m-1}\tau^{m-1}\right),\ \forall\, m\in\mathbb{N}^*,\ i=\overline{1,q}$$

Theorem 4. *Let* $f\in C(\mathbb{R}\times\mathbb{R}_+,\mathbb{R})$. *Suppose that:*
(i) exist $\omega>0$ *such that:* $f(t+\omega,u)=f(t,u),\ \forall\ (t,u)\in\mathbb{R}\times\mathbb{R}_+$;
(ii)exists $L,\gamma>0$ *such that:* $|f(t,u)-f(t,v)|\leqslant L|u-v|,\ \forall\, t\in\mathbb{R},\ \forall\, u,v\in\mathbb{R}_+$
$|f(t_1,u)-f(t_2,u)|\leqslant\gamma|t_1-t_2|,\ \forall\, t_1,t_2\in\mathbb{R},\ \forall\, u\in\mathbb{R}_+$;
(iii) exist $M\geqslant 0$ *such that* $|f(t,u)|\leqslant M,\ \forall\, t\in\mathbb{R},\ \forall\, u\in\mathbb{R}_+$;
(iv) $L\tau<1$. *Then, the solution of the integral equation has unique solution approximated with the error estimation:*

$$(9)\ \left\|x^*(t_i)-\overline{x_m(t_i)}\right\|\leqslant\frac{L\tau^m}{1-L\tau}M\tau+\frac{\tau^2}{4n}\cdot\frac{\gamma+2ML}{1-L\tau},\ m\in\mathbb{N}^*,\ i=\overline{0,q}\ .$$

Proof. By the Banach's principle and Picard's Theorem we have:

$$\left|x^*(t_i)-\overline{x_m(t_i)}\right| = \left|x^*(t_i)-x_m(t_i)+x_m(t_i)-\overline{x_m(t_i)}\right|\leqslant$$

$$\leqslant\frac{L\tau^m}{1-L\tau}|x_0-x_1|+\frac{\tau^2}{4n}(\gamma+2ML)\cdot\frac{1-L^m\tau^m}{1-L\tau}$$

□

References

[1] A. Bica, "The error estimation in terms of the first derivative in a numerical method for the solution of delay integral equation from biomathematics," *Revue d'analyse numerique et de theorie de l'approximatin*, Vol. 34 (1), pp. 23-36, 2005.

[2] P. Cerone, S. S. Dragomir, *Trapezoidal and midpoint-type rules from inequalities point of view,* Handbook of analytic computational methods in applied mathematics (G. Anastassiou ed.), Chapman and Hall/CRC, New York, 2000.

[3] K. L. Cooke, J. L. Kaplan "A periodicity threshold theorem for epidemics and population growth," *Math. Biosciences*, Vol. 31, pp. 87-104, 1976.

[4] D. Guo, V. Lakshmikanthan "Positive solutions of nonlinear integral equations arising infectious diseases," *J. Math. Anal. Appl.*, Vol. 134, pp. 1-8, 1988.

[5] C. Iancu "A numerical method for approximating the solution of an integral equation from biomathematics," *Studia Univ. Babes-Bolyai, Mathematica*, Vol. 43 (4), pp. 37-45, 1988.

[6] I. A. Rus, *Principles and aplications of fixed point theory,* Ed.Dacia, Cluj-Napoca, 1979, (in Romanian).

[7] I. A. Rus, "Fiber Picard operators on generalized metric spaces and an application," *Scripta Scientiarum Mathematicarum*, Vol. 1, Facs.II, pp. 355-363, 1989.

[8] L. R. Williams, R. W. Leggett, "Nonzero solutions of nonlinear integral equations modelling infectious disease," *SIAM J. Math. Anal.*, Vol. 13, pp. 112-121, 1982.

Loredana-Florentina Galea
Agora University of Oradea
Faculty of Law and Economics
Piata Tineretului nr.8, 410526, Oradea, Romania
E-mail: loredana.galea@univagora.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 304-309

Protensity in Agent-Oriented Software. Role, Paths, Example

Alexandru V. Georgescu, Alina E. Lascu, Boldur E. Bărbat

Abstract: Remarking that applied research in the Semantic Web area focuses on tools expressed in formal specifications (i.e., on *syntax* and *semantics*), the paper asserts the need to consider *pragmatics* too, as the only component involving directly the user (through the process of *semiosis*). Since in most intellectual activities users could be assisted by agents, the target is to develop affordable (i.e., purely software) agents able to interpret, evaluate and/or process protensional information contained in multimodal (predominantly sound-based) messages; the agents should have a powerful temporal dimension and should be conceptualised as "Virtual *x*", where *x* stays for various music-related activities. In approaching the target, the paper has four objectives: a) defending the rationale of the undertaking; b) explaining the new concepts proposed and outlining the approach; c) setting out the design space for the first "protensional agents"); d) assessing the approach via a toy problem (designing a virtual disc jockey). The paper concludes that: Computer-Aided Semiosis is a promising field of agent-oriented research also for sound-based messages; protensity should be investigated for both music-related activities and the temporal dimension of bodiless agents; the solution given to the toy problem indicates that the approach is workable and could be applied in an experimental model of a virtual guitar teacher in 2008.
Keywords: Protensity; Computer-Aided Semiosis (CAS); Protensional Agents (PA) as e-Maieuts; Agent-Oriented Software (AOS); Virtual Disc Jockey (VDJ).

1 Introduction

Recently Tim Berners-Lee coined the term GGG (Giant Global Graph) to explain - not to replace - the well-known "*Semantic Web*" in historical evolution, starting from its first stage III (International Information Infrastructure (timbl's blog: http://dig.csail.mit.edu/breadcrumbs/node/215): "The realization was, "It isn't the cables, it is the computers which are interesting". [At *WWW* stage] "It isn't the computers, but the documents" [...] now, "It's not the documents, it is the things they are about which are *important*"". In line with the dictionaries, *importance* means (also) "*significance or prominence*". In short, what it means for the human user, in the given context. On the other hand, "*The Semantic Web effort* (http://www.w3.org/sw) provides standards and technologies for the definition and exchange of metadata and ontologies. Available standard proposals provide ways to define the syntax (RDF) and semantics of metadata based on ontologies (OWL)" [11]. Likewise, the "*Social Semantic Desktop*" concept is "very much related to the Semantic Web but is distinct insofar its main concern is the personal use of information. [...] Ontologies allow the user to express personal mental models and form the semantic glue interconnecting information and systems" (http://en.wikipedia.org/wiki/Semantic_desktop). Thus, the user as the very raison d'être of the Semantic Web seems forgotten: the impressing technologies provided so far concern *syntax* and *semantics*, i.e. the epistematic (explicitly algorithmic) aspects of knowledge processing made possible by those technologies. Nothing about the third element of Peirce's semiotic triad (of sign, object and interpretant). Indeed, only the relationships between signs and signs (*syntax*) or between signs and objects (*semantics*) are considered, while the crucial relationship between signs and interpretants (*pragmatics*) is disregarded, albeit it is the only that involves directly the users - through the process of *semiosis* - and is the only issue that matters for them.

Since in most intellectual activities users could be assisted by agents, Computer-Aided Semiosis (CAS) is ever more relevant. However, the approach in [7] needs two basic steps ahead: to consider also *sounds* (avoiding graphocentrism [10]) and to deal with *music* (avoiding logocentrism too [10]). Considering the context of academic research (above all, its restrictions [5, 8]) the target is: to develop affordable (i.e., purely software) agents able to interpret, evaluate and/or process protensional information contained in multimodal (predominantly sound-based) messages; the agents should have a powerful temporal dimension, should be inherently action-oriented and highly personalised; they are conceptualised as "Virtual *x*", where *x* stays for various activities linked to music. Those that will reach the stage of experimental model, should be carried out through software entities - called "Protensional Agents" (PA) - interacting with the user as interface agents and represented on the screen as pseudoavatars.

As regards this paper, the complex target (to be reached after three years within a PhD thesis [14]) is pursued through four steps (objectives): a) defending the rationale of the undertaking; b) explaining the new concepts proposed and outlining the approach; c) setting out the design space for the first PA); d) assessing the approach

via a toy problem (designing a VDJ). Thus, after presenting *related work* and *history* (Section 2), the first two objectives are dealt with *explaining the title* (Section 3). On this groundwork the last two objectives are worked at, outlining an initial *design space for protensional agents and instantiating it for a VDJ* (Section 4). Conclusions and future work close the paper (Section 5).

2 Related Work and History

Since this undertaking is unconventional - both *protensity* and *semiosis* just enter artificial intelligence - *related work* referred to here regards merely some relevant approaches to VDJ, revealing the main trends in the area while *history* abridges the paper CAS-background and links to related projects.

VDJs on the Web. The term "Virtual Disk Jockey" denotes rather a loosely labelled set of software products designed specifically to help humans in their "DJ-ing" activity, than a software category as such. Most VDJs, like "Mixxx"(http://mixxx.sourceforge.net/), embody a user-friendly interface for choosing and fine-tuning music from existing collections. However, some of them, as "Virtual DJ" (http://www.virtualdj.com/) or "Virtual DJ Studio" (http://www.vdj.net/) provide, beside devices for simple mixing activities (standard controls, volume control, pitch control, equalizers), features coming a bit closer to intelligent help like recognition of music style. Also, beat detection/matching is a must for VDJs (all of the mentioned VDJs have this feature) and is a step toward interpreting sound based messages. A rare implementation of VDJ in line with the architecture outlined in Section 4 is shown in [20]: the disc jockey makes decisions about the next piece of music depending on the success of the previous one (observing the audience through a TV camera). An interesting trend seems to be the shift towards a kind of "virtual potpourri composer" as the "Data jockey" described in [17]: it is a kind of VDJ that aides the users in finding interesting juxtapositions between pieces of music stored on their computers based on different music "attributes" (harmonicity, average brightness, bpm, etc), this way, gaining new insights on their music.

CAS *History* begins within [7], where the term was launched from an anthropocentric perspective: due to multimodal interfaces, computers could assist humans in understanding (above all, trans-cultural) messages, lessening linguistic hurdles (as the "traduttore-traditore" effect), the logocratic pressure of (spoken or written) text, response-time criticality, as well as the danger of distortions and noise, via a major upgrade in communication granularity: (one) idea instead of (many) *words*. Despite dealing only with extensity-related aspects (e.g., attaching semantic value to the iconic space), a first step towards protensity-related aspects was suggesting the anisotropy of the iconic space. (Besides, from Eco's point of view "semiotics is concerned with everything that can be taken as a sign [...], signs take the form of words, images, sounds, gestures and objects" [12], hence, from a CAS stance, extensity of icon-based messages and protensity of sound-based ones could be investigated alike.)

3 Rationale and Approach. Explaining the Title

"*Protensity*" is examined as investigation topic in itself and as newcomer in IT applied research. Since nowadays explaining "*Agent-Oriented Software*" is vain, the focus is on "*in*", confining the research area and showing the links. The rest is self-explaining: "*Role*" filters the rationale through the sieve of the paper objectives; "*Paths*" means a two-way approach: top-down (from concepts to design space) and bottom-up (from VDJ to more complex "Virtual *x*"); "*Example*" refers to an operational, albeit simple, VDJ.

Protensity. This concept was always a poor relative in the family of time-related features. Said to have been invented to match the space-related feature of extensity, it was defined as "the temporal dimension of consciousness" [9] or "the attribute of a mental process characterised by its temporality or movement forward in time" (cancerweb.ncl.ac.uk/cgi-bin/omd?protensity). More focused, it is "the subjective experience of time, as distinguished from physical (clock) time" (http://sonify.psych.gatech.edu/ walkerb/ classes/perception/pdf/07-time.doc.pdf). "Unlike many areas of research, the investigation of subjective time, or protensity, has been conducted from the perspective of a number of different disciplines (e.g., philosophical, biological)" [22]. Albeit various perspectives could be useful in future work, here is relevant the psychophysical stance: "What we ordinarily call a single sensation has not only a characteristic quality but it is also quantitatively determined in respect of intensity, protensity (or duration) and extensity [...] there is an element in our concrete time-perception which has no place in our abstract conception of time" (http://www. 1911encyclopedia.org/LoveToKnow_1911).

As a result, protensity is studied in diverse contexts but oversimplified or very limited; for instance:

– *Time-based or dynamic media* (audio, video, music, choreography), having "basic, time-related attributes like a duration or a starting time" [1].

– *Phonology*, where protensity attributes "are only represented by the feature tense/lax"(www.apa.org/ divisions/div21/Meetings/SymAbstracts2002.pdf).

– *The influence of stress* on subjective time (crucial for military purposes) [22].

– *Music education*, where Ortmann's "seminal 1922 article, "The Sensorial Basis of Music Appreciation," [...] has received little attention. [...] What should we teach when we teach music? [...] In coining transtensity as the character of being extended across something, Ortmann [...] reasoned that transtensity, intensity, and protensity are primary since, if any one is eliminated, the entire sensation is eliminated" [16].

Hence, investigating protensity has three strands of motivation:

a) Promising research *per se*: a1) Here and now, "Prigogine's idea that the most interesting scientific activities seem to occur at domain interfaces" [5] shows the only affordable path for applied research. a2) If computers have to deal with meanings, they should be first able to assist end-users in a basic process that was until now an exclusive human attribute: *semiosis*. Indeed, through the process of semiosis, the message receiver "fills the message with significance" [12]. a3) If atemporal text or images can be accepted as messages where computers could assist humans in understanding them, why temporal sound-based messages could not be treated likewise? (from a technologic-historical perspective it should be easier: after all, radio is almost a half-century older than television). a4) Whereas extensity was (more or less) examined - albeit rather implicitly, e.g., in [7] -, protensity was almost ignored. a5) Because of the unidirectionality of time, uncertainty is much more important in time than it is in space.

b) Immediate applicative potential: b1) The existence of so many VDJ (see previous section) shows two aspects: the presence of a genuine and seemingly powerful social command; the syncretic stage of applied research in this subfield. b2) The large application area offered by the Semantic Web (see Section 1). b3) Time-sensitive software is not anymore difficult to develop, due to agent technology [6, 13].

c) Ancillary research, essential for other agent-oriented undertakings (major motivation for this paper): c1) Non-algorithmic e-Learning (an approach based on eMaieutics is developed in a related paper). c2) CAS and trans-cultural communication [7]. c3) Gödelian self-reference for emulating agent self-awareness [4, 6, 19]. c4) Emergence, complexity and time [18, 19]. c5) Uncertainty due to future contingents [8, 3].

Agent-Oriented Software. *Music* and *agents* are linked through their intrinsic *process* nature. Obvious in practice, the acceptance of their temporal dimension penetrates now theory too: "my personal view is more along the lines of process philosophy which would regard music not as an object, but as an emerging event" [21]; the agent is defined as process in a standard [13]. Thus, both *semiosis* and *protensity* become unavoidable in agent-oriented research. However, while the link is obvious in the first two endeavours, for the others some explanation is needed: c3) As regards self-awareness: "somatoception, the awareness of one's own body, involves many specialized sensors" and "at this very basic level, self-representation is bodily-representation, and the self is known as, and in terms of, its body" [2]; hence, for bodiless agents "the expected emergence of a primitive "I" should be catalysed through powerful temporal dimension" [4]. c4) Protensity could help in emulating emergence, via investigating non-uniform time perception [19, 13]: "In time conceived as physical there is no trace of intensity; in time psychically experienced duration is primarily an intensive magnitude, witness the comparison of times when we are bored with others when we are amused." (http://www.1911encyclopedia.org/LoveToKnow_1911). c5) Since time is a paramount resource in both real world and modern IT, decision-making under uncertainty becomes vital; hence "uncertain semiosis", too.

4 Protensional Agent Design-Space. Instantiating it for VDJ

Considering the need for modularity, flexibility, and stepwise development (based on successive prototyping, with a pace depending on the progress of the experimental models in the projects involved), as well as the research plan for 2008 (when all PA will address e-Learning environments), the design space for PA is a restricted subset of the Cartesian product: $\mathbf{S}_{PA} = \mathbf{S}_{Socrates} \times \mathbf{S}_{CAS}$, where $S_{Socrates}$ (the design space for virtual maieuts) is defined in the related paper as: $\mathbf{S}_{Socrates} = \mathbf{S}_{Maieutics} \times \mathbf{S}_{Agents}$.

On the other hand, S_{Agents} is set up depending on the model for the application in sight. For instance, a virtual guitar teacher, will be less complex than a virtual musicologist but will need much more maieutic features than the other, while a VDJ could lack even some weak agency features (e.g., since it talks only to humans, its communication skills could be very limited); likewise, a VDJ does not need maieutic design-space dimensions - though it will need a primitive ontology (to chose the next piece of music from the available ones) and exception mechanisms (to react promptly to audience stimuli).

As regards S_{CAS}, except the temporal dimension and a kind of music sub-space, in line with [21], depending on the physical characteristics of sound that are involved at the level of protensity-based semiosis performed by the model on hand, it is too early to comment upon (the VDJ just *follows* audience preferences without *influencing* them).

Smart DJ, the toy problem of this undertaking (illustrated in Fig. 1), is a VDJ that can automatically choose the songs by "hearing" and interpreting the audience's reaction to the type of music previously played. It uses a microphone to "listen" to the sounds/noises that are emitted by the audience/listeners and tries to determine if they *mean* acceptance (applauses, shrieks, whistles) or rejection (hoots, etc.) of the currently played song. Also, changes in the intensity of a certain sound/noise type are tracked, thus different pieces of music can be classified according to some criteria. When a song is accepted or rejected, Smart DJ analyses its attributes to find out the cause (genre, artist, BPM, harmonicity or other attributes). This analysis can be done for example, by playing another song that has one or more identical attributes and observing the audience's reaction. Of course, the lack of any reaction (considering in this context, total or partial silence) must be interpreted.

Figure 1: Interface of Smart DJ

Smart DJ is not entirely automatic because the user can intervene any time in selecting the next song (it also can be used as a normal VDJ, in *manual mode*, where the users select all the songs). The user's intervention is taken as a *hint* vis-a-vis the music type (genre, artist, etc.) to be played or the kind of music to substitute the rejected one. Of course, *the accent* is on correctly interpreting sound/noise *meaning* not on accurately analysing the spectrum of the data that comes through the microphone. This might lead to certain "confusions" of sound sources (a door creak may be taken as a whistle, or something like a ship horn, as a hoot).

Smart DJ'*s user interface* (Fig. 1) simulates a straightforward DJ mixing console interface with the two well known disk slots, controls for equalization, pitch, track and synchronization for each disk, peak meters and headset

redirect button for each disk, spectrum display with beat information, overall volume control and the *crossfader*. Also it has a view of the audio files form the local and remote disks which can be categorised by genre, BPM or other categories.

What makes it different from other VDJs is the *Automatic button* which toggles the mechanism of automatically selecting musical pieces described above.

NOTE: Currently, Smart DJ is still under development and some modifications in the interface may occur but the general layout will remain. An avatar-like user interface is also considered as future work, this way this VDJ will look more like an interface agent as it was mentioned in the first section.

5 Conclusions and Future Work

Due to the two-way approach, the assessment is based on two kind of conclusions: A) top-down (from concepts to design space) and B) bottom-up (from VDJ to more complex "Virtual *x*")

A1. Computer-Aided Semiosis is a promising field of agent-oriented research also for sound-based messages.

A2. Protensity as message attribute is a motivating research topic both *per se* (in particular, for music-related activities) and as test bench for agent-oriented software (mainly for investigating the temporal dimension of bodiless agents).

A3. The partial analogy with extensity is an affordable starting point for exploring protensity.

A4. The proposed design space suits the development of diverse protensional agents.

B1. The toy VDJ model presented shows that the approach is both workable and affordable with scarce resources because - albeit quite simple - the VDJ reacts in a relevant manner to audience stimuli.

B2. On this groundwork, a virtual guitar teacher could reach the stage of experimental model in 2008.

B3. The VDJ architectonics is a fitting stem cell for designing test settings for self-referencing agents.

Future work. Beside the virtual guitar teacher, a middle-range target is to develop a self-referencing protensional agent with a primeval sense of time, expressed by the capacity to assess its speed of learning and to clone itself into a smarter agent according to its own evaluation. Long-range targets are in line with the roadmap set up in [14].

Acknowledgment. This work was supported by the Ministry of Education and Research through Grant CNCSIS 33/2007.

References

[1] J. C. Anderson, Musical Identity, *The Journal of Aesthetics and Art Criticism*, 40, 3, 285-291,1982.

[2] M. L. Anderson, D. R. Perlis, The roots of self-awareness. *Phenomenology and the Cognitive Sciences*, 4, 297-333, Springer, 2005.

[3] B. E. Bărbat, DOMINO: Trivalent Logic Semantics in Bivalent Syntax Clothes. *International Journal of Computers, Communications and Control*, 2, 4, 303-313, 2007.

[4] B. E. Bărbat, A. Moiceanu. I, Agent. *The good, the bad and the unexpected: The user and the future of information and communication technologies* (B. Sapio et al, Eds.), Conf. Proc., Brussels, COST Action 298 Participation in the Broadband Society, CD-ROM ISBN: 5-901907-17-5, 2007.

[5] B. E. Bărbat, A. Moiceanu, S. Pleşca, S. C. Negulescu. Affordability and Paradigms in Agent-Based Systems. *Computer Science Journal of Moldova*, 15, 2(44), 178-195, 2007.

[6] B. E. Bărbat, A. Moiceanu, I. Pah. Gödelian Self-Reference in Agent-Oriented Software. *Proc. of the 11th WSEAS International Conference on COMPUTERS (ICCOMP '07)* (N.E. Mastorakis et al, Eds.), 92-97, Agios Nikolaos, Crete, 2007.

[7] B. E. Bărbat, S. C. Negulescu, A. E. Lascu, E. M. Popa. Computer-Aided Semiosis. Threads, Trends, Threats. *Proc. of the 11th WSEAS International Conference on COMPUTERS (ICCOMP '07)* (N.E. Mastorakis et al, Eds.), 269-274, Agios Nikolaos, Crete, 2007.

[8] B. E. Bărbat, S. C. Negulescu, S. Pleşca. Emergence as Leverage and Non-Algorithmic Approaches in Agent-Oriented Software. *Studies in Informatics and Control Journal*, 16, 4, 321-332, 2007.

[9] E. G. Boring, The Physical Dimensions of Consciousness. Century Co, New York, 1933.

[10] D. Chandler, *Biases of the Ear and Eye*. http://www.aber.ac.uk/media/Documents/litoral/litoral.html [Accessed, 12/8/2007], WWW document, 2000.

[11] S. Decker, M. Frank. *The Social Semantic Desktop*. Digital Enterprise Research Institute (DERI) Technical Report 2004-05-02, 2004

[12] U. Eco, *The Limits of Interpretation*, Bloomington, Indiana University Press, 2005.

[13] FIPA TC Agent Management. FIPA *Agent Management Specification*. Standard SC00023K (2004/18/03). http://www.fipa.org/specs/fipa00023/ SC00023K.pdf, 2004.

[14] A. V. Georgescu, *Agent-Oriented Semiosis for Protensity Messages. Application in Musicology.* (PhD Thesis in preparation.)

[15] A. V. Georgescu, Protensional Agent as Virtual Guitar Teacher. (In preparation.)

[16] D. J. Gonzol, *Otto Rudolph Ortmann, Music Philosophy, and Music Education*. Philosophy of Music Education Review, 12, 2, 160-180, 2004.

[17] A. Norman, X. Amatriain, Data jockey, a tool for meta-data enhanced digital djing and active listening. *Proceedings of the International Computer Music Conference*, Copenhagen, 2007.

[18] S. Pleşca, *Emergence in Multi-Agent Systems. Application in Stigmergic Coordination* (PhD Thesis in preparation.)

[19] S. Pleşca, A. Moiceanu, E.M. Popa. Self-Referencing Agents in Non-Algorithmic e-Learning. (In preparation.)

[20] M. Singh, Virtual DJ. Msc In Distributed Multimedia Systems, 2001.

[21] J. Steyn, Introducing Music Space. www.interactivemusicnetwork.org/events/ Fourth_OpenWorkshop_2004/musicspace.html, 2004

[22] J. L. Weaver, et al. The Influence of Stress and Individual Differences on Subjective Time. *Abstract Proc. for Midyear Symp. Contemporary and Emerging Issues in Human Factors, Engineering, and Military Psychology*, 2002.

Alexandru V. Georgescu
"Politehnica" University of Timişoara
Vasile Pârvan Blvd. 2, 300223, Timişoara, Romania
E-mail: alexandrugeorgescu@gmail.com

Alina E. Lascu*, Boldur E. Bărbat**
"Lucian Blaga" University of Sibiu
*Faculty of Political Sciences, International Relations and Security Studies
**Faculty of Sciences
Ion Raţiu St. 5-7, 550012, Sibiu, Romania
E-mail: *alina.lascu@gmail.com, **bbarbat@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 310-315

Computer Study of Some Dynamical Nonlinear Optical Systems

Mihaela Ghelmez, Valerica Ninulescu

Abstract: This paper presents some computer studies of the behavior of dynamical systems in nonlinear optics. These studies were realized together by students and professors as student homework, and presented at the student annual Scientific Session. They were awarded with prizes and diplomas. The Lorenz model, the logistic equation, Hénon and Ikeda models were used for simulating the chaotic behavior of the systems in QBASIC. Then systems able to generate practical test-functions (defined as functions which differ to zero on a certain interval and possessing only a finite number of continuous derivatives on the whole real axis) are studied. The shape of the output signal, obtained by numerical simulations in Matlab based on Runge-Kutta functions, is analyzed, being shown that for high-frequency inputs an external observer could notice, in certain condition, the generation of two different pulses corresponding to two distinct envelopes. Results are in accordance with other for characterizing the nonlinear optical and dielectric materials.
Keywords: dynamical systems, nonlinear optics, chaotic behavior, Runge-Kutta functions, test functions

1 Introduction

Taking into account changes in the Physics curricula and the actual number of hours allotted for the physics course and applications, students are asked to complete their instruction in this matter by performed some home work. These works are realized with the help of different computer programs, and were also successfully presented at the student annual scientific session. This paper presents some results obtained last academic year at "Politehnica" University, and awarded with prizes and diplomas.

2 Study of the dynamical systems theory and applications

One direction in our activity with students is the study of the dynamical systems theory and its applications to nonlinear optics. Students are initiated for using paradigm systems, for simulating the chaotic bistable behavior of some systems: Lorenz model of turbulence, logistic equation, Hénon two-dimensional discrete equation [1]. The Lorenz model, consisting of three ordinary nonlinear differential equations:

$$x' = 10(-x+y)$$
$$y' = -y+(28-z)x$$
$$z' = -(8/3)z+xy$$

(1)

was used for predicting the evolution of the system, the result of calculus being presented in Fig.1. For the sake of simplicity, the Euler method was applied. In order to be close to the attractor, a great number of iterations were plotted. The logistic equation:

$$x_{n+1} = 4ax_n(1+x_n)$$
$$x_0 \in [0,1]$$
$$a \in [0,1]$$

(2)

is the simplest smooth equation that exhibits a large amount of complex dynamics. The bifurcation diagram (Fig.2), i.e. a plot of the iterated values $x_n n > N$ (N large enough and x_0 a random value in (0,1) interval) versus the control parameter a, reveals a lot of interesting phenomena: pitch fork bifurcations, onset of chaos, noise bands merge, and narrow windows of periodic evolution.

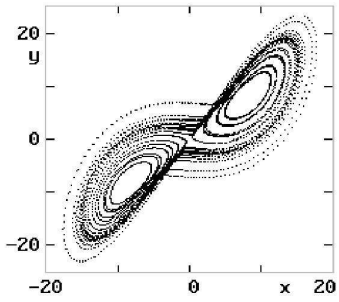


Figure 1: Lorenz strange attractor

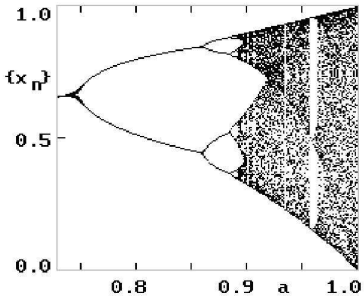


Figure 2: Bifurcation and chaos in logistic equation

A similar diagram is obtained with the Hénon model:

$$x_{n+1} = y_n + 1 - Ax_n^2$$
$$y_{n+1} = 0.3x_n$$

(3)

when the plot x_n versus A is represented and $A \in ([0,2]$. A simple model, which exhibits a complex behavior and gives rise to optical turbulence, is the Ikeda model of optical bistability [2]. Suppose a laser beam is injected into a ring cavity containing a cell with two level atoms in the dispersive limit, the transfer characteristics is a difference equation with the form:

$$E_{n+1} + A + BE_n \exp[i|E_n|^2 - \delta]$$

(4)

Here E_n and A are proportional to the transmitted electric field of the laser and the constant input electric field, respectively; B and δ are some constants. The stationary solution:

$$A = |E|[1 + b^2 - 2B\cos(E^2 - \delta)]^{1/2}$$

(5)

where A is a real quantity, $B = 0.5$, and $\delta = 0$, is represented in Fig.3. It is remarquable the multivalued response of the cavity to a constant incident light.

A similar representation was made for the modulus of the output electric field of the laser signal in Fig.4. We have considered the set of initial conditions [(1,1),(2.5,0.25), (1.9,1.9)] 150 unplotted iterations, and 20 plotted iterations. The stationary solution is represented as a dotted line. Stable stationary states are reaching during the evolution (solid curves). Some bifurcations of these states, onset of chaos, periodic windows, reverse bifurcations to a stable stationary state, the co-existence of two attractors of period two on the first branch of the $|E| - A$ diagram, give the complex dynamics of the system.

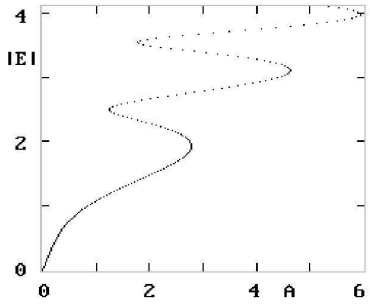


Figure 3: Stationary states curve for the Ikeda model of optical bistability

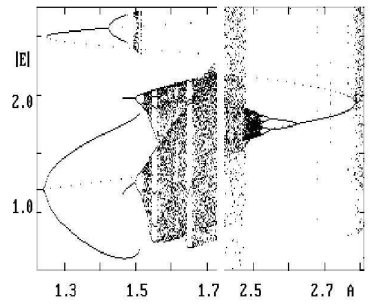


Figure 4: Ikeda model. Bifurcation diagram. Dotted line represents the unstable stationary states. When $A < 1.225$, stationary states are stable.

Fig.5 presents the asymptotic evolution in (E) complex plane, for $A = 2.0$, when a chaotic dynamics is installed. ($E_0 = (1.1)$ like initial value of E). The enlargement of the indicated rectangular region suggests the strange attractor dynamic evolution.

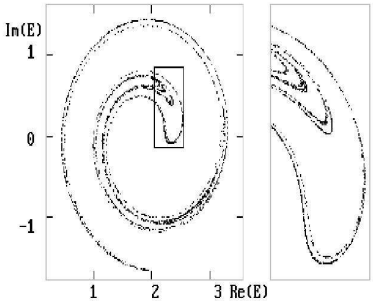


Figure 5: Ikeda attractor in the E complex plane for $A = 2.0$, and a detail showing the self-similar structure, a specific feature of a strange attractor

The programme is performed in QBASIC and finally some options can be chosen by the student, concerning the values domain, without modifying of this programme: new plots occur in terms of these new values and one can see immediately the influence on the dynamic regime of the system. Graphical images can be wholly displayed, but their realization like a numerical experiment in the student's presence is very instructive. Observing a bad contrast for the images from Fig.4 and 5, a simple command can modify the "positive" version into a negative and more suggestive one.

3 Computer study of material answer to some external signals

We are looking for a mathematical model for describing the dynamics of phenomena taking place inside a material [3], under the influence of external optical pulses. Due to the high optical frequency, we cannot use a linear equation of evolution, which answer would be close to zero. We begin our study by looking for a nonlinear equation of evolution, this nonlinear dynamics being able to generate pulses similar to test-functions. These functions, similar to a Dirac pulse, can be written under the form:

$$\varphi = \exp(\frac{a^2}{\tau^2 - 1})$$

(6)

where $t = t - t_{sym}$, t_{sym} being the middle of the working period. Such a function has nonzero values only for $t \in [-1,1]$. We are looking for a differential equation, which can have as a solution the function φ. However, such an equation cannot generate the test function φ. The existence of such an equation of evolution, beginning to act at an initial moment of time, would involve the necessity for a derivative of certain order n - noted $f^(n)$ to make a "jump" at the initial moment, from the "zero" value to another value which differs to zero. This is in contradiction with the property of the test-functions to have continuous derivatives of any order on the whole real axis (in this case represented by the axis of time). It results that an ideal test-function cannot be generated by a differential equation, but it is quite possible for such an equation to possess as a solution a "practical" test function f, i.e. a function with nonzero values on the interval $t \in [-1,1]$, and a certain number of continuous derivatives on the whole time axis. Therefore, we try to study those evolutions depending only on the values $f, f^{(1)}, \ldots, f^{(n)}$, these values being equal to the values of $\varphi, \varphi^{(1)}, \ldots, \varphi^{(n)}$ at a certain time moment, very close to the initial moment $t = -1$. By using the equations [4]:

$$f^{(1)} = [-2\tau/(\tau^2 - 1)]f$$

$$f^{(2)} = [(6\tau^4 - 2)/(\tau^2 - 1)]f$$

$$f^{(2)}(\tau) = [(0.6\tau^4 - 0.36\tau^2 - 0.2)/(\tau^2 - 1)^4]f(\tau)$$

(7)

and with different initial conditions for $f, f^{(1)}$ and using Runge-Kutta functions in MATLAB, it resulted that the equation that could lead to some functions similar to a rectangular unitary pulse is the last one, since the amplitude is close to unity for more than 2/3 of the integration period). The previous equations can be generalized using computer methods presented in [5].

However, for our mathematical model we have to change these differential equations, because all changes inside the material appear due to the external optical pulse. Therefore, we must first consider null initial conditions for the system, and we must also add a free term in the differential equation - corresponding to the magnitude of the electrical field of the external signal, having a frequency of about 1015 Hz. The working period was chosen approximately equal to the period when the optical signal is received by the detector (about 0.2ms). The differential equation can be written as:

$$f^{(2)}(\tau) = [(0.6\tau^4 - 0.36\tau^2 - 0.2)/(\tau^2 - 1)^4]f(\tau) + u(\tau)$$

(8)

where u is represented by an alternating function with a frequency 10^{11} times greater than the working period of 0.2 ms. By numerical simulations in MATLAB with Runge-Kutta functions, we have obtained for f the results presented in Fig.6 for $u = sin(10^{11}\pi t)$, and in Fig.7. for $u = cos(10^{11}\pi t)$.

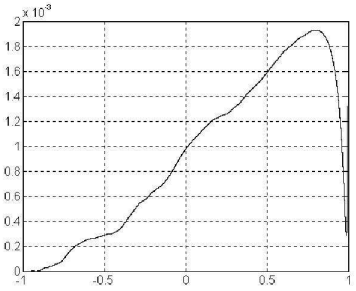


Figure 6: f versus t for $u = sin(10^{11}\pi t)$

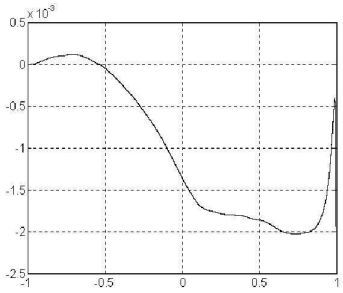


Figure 7: f versus t for $u = cos(10^{11}\pi t)$

The function f, generated by the material under the influence of the external optical pulse, can be integrated on this working time, the result of this operation representing the physical quantity measured by the external observer. It can be noticed that, for the external observer, the behavior of the material presents a slowly varying evolution in time-a single oscillation on the whole working interval-, even thou the input signal is a fast varying one (at an optical frequency), similar to the behavior noticed during the experiments. This allows us to consider our method, based on the use of systems described by differential equations, able to generate pulses similar to test functions, was appropriate. In this case, we are using a similar way for analyzing the behavior of the materials in external electric field, in correlation with the previous results. For simulating the dynamics of the dielectric properties of the material under the influence of an external electrical pulse [6], supposed to be constant over a certain time interval, we will consider a differential equation, able to generate a practical test-function similar to a Dirac pulse, and having also a free term u(t), corresponding to the magnitude of the external signal. Considering as origin of time the middle of the working interval, we chose the following equation:

$$\varphi^{(2)} = [(6\tau^4 - 2)/(\tau^2 - 1)^4]\varphi + u(\tau)$$

(9)

where $u(t) = 1$ on the working interval (-0.99; 0.99). This choise is justified by its sharpening, easier to be observed graphically. The observer "see" the function f, generated inside the material, under the form of a kind of "progressive" wave, with a time variable phase on the working range. The influence of the function φ is received like a sum of all effects, which can be represented as an integral of function φ multiplied by the progressive wave. Let by a quantity "z", corresponding to the dielectric properties of the material. This could be written as:

$$z^{(1)}(\tau) = \varphi(\tau)sin(\pi t - \Phi)$$

(10)

where Φ represents an initial phase of the progressive wave, caused by an internal response inside the material. The function $z(\tau)$ is represented in Figure 8, for $u(\tau) = 1$ and $\Phi = \pi/12$. It can be noticed that the function z is equal to zero at the initial moment, and then presents a minimum value at a moment close to the initial one, before ending its evolution at a certain final value, that models the dielectric experimental behavior of the materials with a weak conduction [7], presented in Fig. 8

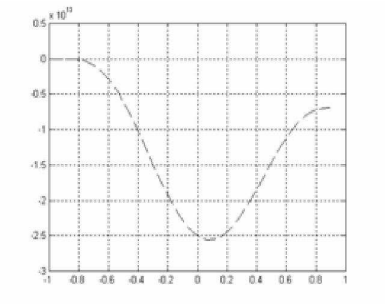


Figure 8: The dielectric experimental behavior of the materials with a weak conduction

4 Summary and Conclusions

Computers use at Politehnica University of Bucharest represents nowadays usual tools for students. The new educational units which started last years at our institute are based on the computer aided engineering education: teaching, learning, evaluation of the knowledge and, in general, the dialog between teachers and students - facilitated by computers. Using computers increasingly makes the students' independent work. In the paper we presented some of the obtained results that were presented at the students Scientific Research Session. Two complementary examples connected with the nonlinear optics study are given. These works are mainly results of the students' research with computers programs. The study underlined the role of the theoretical models in order to estimate the evolution of some dynamical systems, and the possibility to connect the experimental results with the computer study.

References

[1] K.H. Becker, M. Dörfer, *Dynamical Systems and Fractals,* Cambridge University Press,1991.

[2] K.Ikeda *Optics Comm.*, Vol. 30, pp. 257, 1979.

[3] M. Ghelmez (Dumitru), B. Dumitru, A. Sterian, R. I. Trascu, "Experimental and computer studies of the functional activity in laser field of some components of the biological membrane", SPIEVol.4397, Ed.P.Atanasov, S.Kartaleva, pp.385-389, 2001 *SPIE Proc.*, Vol. 4397, pp.385-389, 2001.

[4] F. Doboga, G.Toma, St.Pusca, M.Ghelmez, C.Morarescu, "Filtering Aspects of Practical Test-Functions and the Ergodic Hypothesis", Lecture Notes in Computer Science 3980, p. 563-568, Springer 2005 *LNCS*, Vol. 3482, pp.563-568, 2005.

[5] G.Toma, "Practical Test Functions Generated by computer Algorithms", Lecture Notes in Computer Science 3980, p. 576-584, Springer 2005 *LNCS*, Vol. 3482, pp.563-568, 2005.

[6] V.Tareev, *Physics of Dielectric Materials,* Mir Publisher, Moscow, 1975

[7] M.Ghelmez (Dumitru), Nonlinear Optical Effects in Biological membrane models, *Nonlinear Optical Effects in Biological membrane models,* Ed. Printech, Bucharest, 2005

Mihaela Ghelmez (Dumitru), Valerica Ninulescu
"Politehnica" University of Bucharest
Physics Department
Splaiul Independentei 313, 77 206 Bucharest, Romania E-mail: mghelmez@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 316-321

Dynamically Organization of Educational Contents for E-Learning

Dragana Glušac

Abstract: The current educational system in our region (which part is Serbia) is in a transition period. In this paper will be shortly present two isues: need for E learning in modern society and some theoretical principle of electronic education. The main objectiv is increasing quality and efficiencey of education. This paper explains and systematize the definition and structure for E learning. In order to be able to acomplish it's primary role, the school has to follow the changes in the society, and sometimes even to be ahead. As a significant society institution, the school is daily exposed to diffrent expectations and preasures from both, inside and outside. Each effort to describe an existing state in education is facing the fact that it is a mix of some old patterns of behaviour and some new models of educational practice, which are in the phase of forming. Also, everything is ilustration thrue the project Distance Learning at the our Faculty.

1 Introduction

Fast development of technologies has caused many changes in process of education itself. New standards have been adopted, but problems in realization and application of these standards were unavoidable. Many organizations put great efforts into researching the possibilities of new technologies in process of education.

Here are some of the questions that may appear during the creating of distance learning. These problems have appeared on the occasion of creating models of distance learning as a past of a project "Distance Learning System (DLS) based on Internet technologies using multimedia educational softwares" at the Technical Faculty "Mihajlo Pupin" in Zrenjanin.

The Technical Faculty "Mihajlo Pupin" Zrenjanin began as the Pedagogical Technical Faculty in 1974, as high educational scientific organization for the schooling personnel for the polytechnic and teaching. During 1979 the Faculty educated also the informatics teachers. Dean is the highest managing Faculty organ. The teaching all levels of studies and all profiles and scientific fields is accomplished trough new innovated teaching curriculum on all years of studies. Today, the Faculty are the most important eduactional institution in Banat, and there is the big need for modernisation of educational methodology.

What are the conditions like in our educational system concerning modernization of the teaching process? The Exploration incorporated in the doctoral thesis defended at Technical Faculty Mihajlo Pupin in Zrenjanin [1] showed that nearly 70% of the surveyed teachers was unsatisfied with computer equipment they had, while 10% of the teachers said that the complete teaching of Computer Science was performed without computers.

By analyzing the available documents we came to the similar conclusions. The analyzed documentary source shows that "traditional ways of teaching are predominant in our school system". Specific characteristics of Informatics (as a school subject) didn't contribute to essential changes concerning modernization of teaching methods and the predominant method of teaching in Informatics is still the frontal method of teaching. However, certain schools and teachers that showed successful attempts in modernization of their teaching approach mustn't be neglected. By analyzing a documentary source A we came to the same data: "It is a fact that in our schools a classical approach to a teaching process is still predominant and the main philosophy of teaching is traditional". Such an approach is characteristic, among the others, for traditional methods of teaching. All the mechanisms of our school system which can influence the method of teaching and learning mainly give favour to traditional methods where teaching methods are most commonly used. The point is in the role of teachers - not of students. It has already been said that the predominant teaching method in our schools gives favour to mechanical memory. A strict learning of subject matter is still focused at pure reproduction of the taught contents and not at practical skills and abilities. It often happens that (according to the survey) evaluation is performed on the base of pure contents reproduction in front of the board, and not on computers or other modern equipment. The point is at deficiency of ICT in our schools.

Information literate person understands the role of computers as associates in the process of searching and processing information, but he is equally conscious of the fact that success of the process depends mainly on him and not on the used technology. New, different names for open, flexible and distributed activities in the process of learning and teaching are appearing in this technological and didactic moment: E-learning, Web Based Learning, Web Based Instruction, Internet Based Training, etc.

2 Question of Creating Distance Learning System

At this Faculty exist scientific work, which the most important is Distant Learning Project. Organizational group is Methodic of sciences and educational technology cathedra. The Project was confirmed and financed by Government of Republic Serbia.

A Definition of Distance Education: Distance Education is instructional delivery that does not constrain the student to be physically present in the same location as the instructor. Historically, Distance Education meant correspondence study. Today, audio, video, and computer technologies are more common delivery modes [2].

The term Distance Learning is often interchanged with Distance Education. However, this is inaccurate since institutions/instructors control educational delivery while the student is responsible for learning. In other words, Distance Learning is the result of Distance Education. Another term that has experienced some recent popularity is Distributed Education. This term may represent the trend to utilize a mix of delivery modes for optimal instruction and learning. Distance Education represents a way for connecting and communicating with geographically dispersed individuals and groups. Distance Education defining elements:

1. The separation of teacher and learner during at least a majority of each instructional process.
2. The use of educational media to unite teacher and learner and carry course content.
3. The provision of two-way communication between teacher, tutor, or educational agency and learner.

While designing your distance education program, remember that there are three elements of paramount importance to any successful distance education program:

- Instructional design
- Technology
- Support

This Project has these phases with their activities:

Phase	**Activities**
I Model developing	1. Introduction to existing solutions
	2. Identification of possible solutions
	3. Defining of special knowledge
	4. Team training
	5. Identification of restrictions
	6. Selection of solution
	7. Detailed creation of model
	8. Compliance with new achievements in DL domain
II Concept developing	1. Algorithm for contents defining for model
	2. Making contents for all courses
	3. Testing of concept
	4. Compliance with new achievements in DL domain
III Implementation	1. Resource analysis
	2. Implementation strategy making
	3. System implementation
	4. Maintaining
	5. Results systemizing
	6. Compliance with new achievements in DL domain

The Faculty is mostly equipped with modern equipment. Main issues at the Project is The Possibilities of Distance Education.

One of the answers to the needs and questions above lies in the ability we have to create connections through distance education. By uniting instructor and learner with the use of educational media provided through a mixture of delivery modes, a connection can be established to provide a learning opportunity that would not exist otherwise. This connection can be created without the constraints of time or place, and can utilize a variety of modes that will make it accessible to the learner.

Professional Development Distance Learning Model:

1. Design an online course
2. How to evaluate an online course
3. Distance Education-Related Certificate Programs.

3 Organization and Design an On-line course

Distance Learning, or E-learning, is "a process of knowledge transfer on the WEB by using computer aplications and the system in the studying process". These applications and processes include learning on the web by using computers, digital classrooms, as well as digital collaboration. Contents are transfered over Internet, Intranet, Extranet, audio and video tapes, satellite TV and CD-ROM.

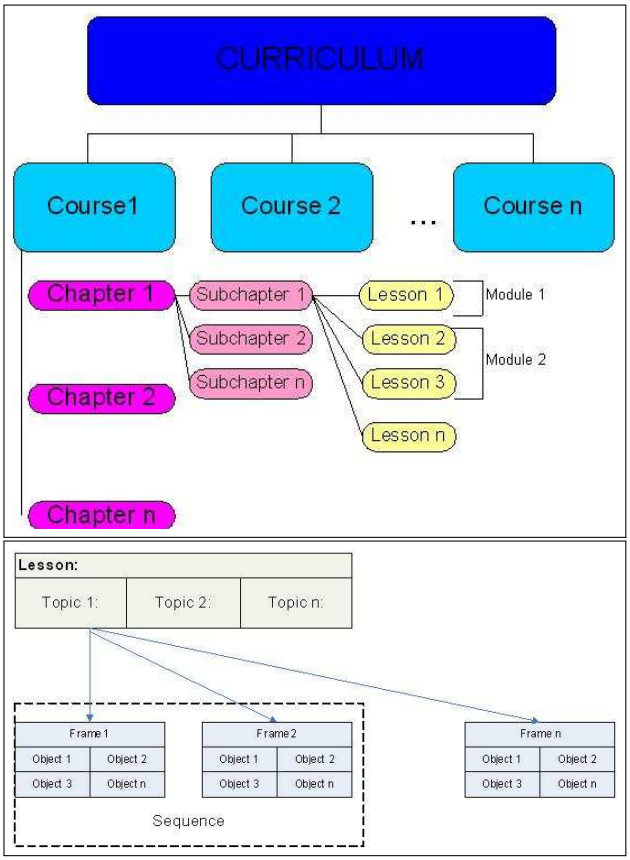


Figure 1: Organization of On-line course

The notion of electronic learning includes a range of fields of WEB intelligence, for example, the use of information systems on the web, ontology engineering, semantic WEB, interaction of man and computers and computer media, managing information on the WEB, searching and finding information and knowledge on the web, web agents, autonomous systems of agents, web mining, as well as, building of new types of applications [3]. WEB pages representing Distance Learning should help students to find necessary information about course, to learn the material and to get to know with the course's object. Well-designed Web pages should urge and help thinking discussions and active participation of students in Distance Learning process..

The elements that must be included in Web pages representing the course are:

- *Basic information about the course and the educator* - the name of the course, educator's working time, information about printed material, review of the course, the rules of evaluating.
- *The communication of the group* - access to the educator's e-mail, discussion group for student-student communication, forms for reporting about problems.

- *Tasks and tests* - distribution of tasks and tests for on-line filling in and submitting, checking the results, tips and tricks, frequently asked questions (FAQ).
- *Material for education* - lectures available in the shape of Web pages and files for downloading.
- *Demonstrations, animations, video, audio* - including the material that can't be presented in classical textual format.
- *Reference material* - The list of additional material in printed or electronic form. These articles should be in public ownership in order to avoid problems with copyrights. In addition, some links to other pages on Internet, which are connected with these themes, similar courses, University's library and other recourses good for completion of the course, may be offered.

The teaching materials could have different characteristics from the ones of traditional sources of information: content is actual and dynamic, content can originate from a primary source, sources can be presented in different ways, information is easy to manipulate with, students can participate on-line, content is available for reading.

Preparation and publishing process of teaching material for E-learning should be organized so that it can be easily found, downloaded and used by users (students). The main problem in the process of implementing E-learning is the lack of compatibility among different platforms: courses developed for a specific system cannot be easily incorporated into similar systems of other producers. A development of contents is a task which demands a great number of supporting means. Even when they are within contents presentation (for example HTML) the adjustment of contents has to be done in order to incorporate the contents in the new platform. In most cases the organization and delivery of contents are tightly connected to the platform's logic. Recently, several solutions for enabling contents exchange have been proposed in order to solve the mentioned problems. Standards on contents structure will make possible the appearance of authorized tools that are independent on the platform but have appropriate advantages both for suppliers and users of educational contents. In other words, in order to transfer a course from one system into the other it is necessary to transfer all elements of that course (lessons, tests, simulations...) along with "metadata". On the other hand, the structure of the starting course has to be maid on the new platform too.

4 Some Examples Of Using Computers In Teaching

There is a lot of research which has shown positive effects of using computers in teaching, especially concerning motivation and quality of acquired knowledge. One of them was a part of Doctoral thesis defended at Technical Faculty "Mihajlo Pupin" in Zrenjanin in 2005, in which motivation of students' who assimilated a specific teaching contents by using computer software was measured.

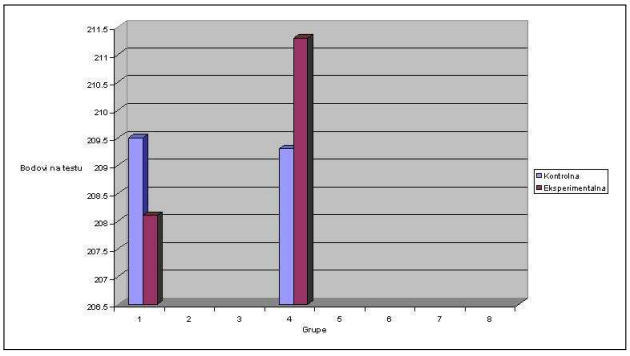
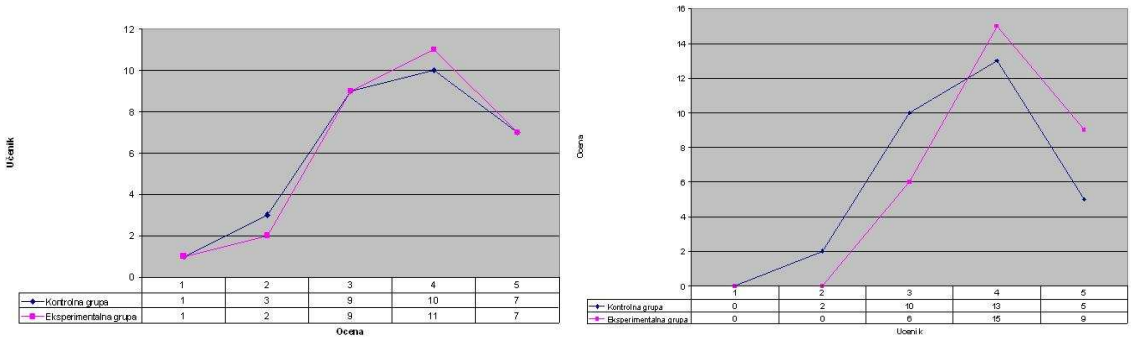


Figure 2: Diagram - Values of the answers in control and experimental group in initial and final achievement measuring

Besides, measuring of the influence of methodic innovations on increase of knowledge level and students' abilities in the field of programming in Pascal, was carried out as well. By comparing the values of answers in control and experimental group in initial and final measuring of achievement motifs it was found out that the experimental group showed progress in the final measuring. By measuring students' achievement motifs the results were got which showed that mid values of the answers were considerably increased in the experimental group in

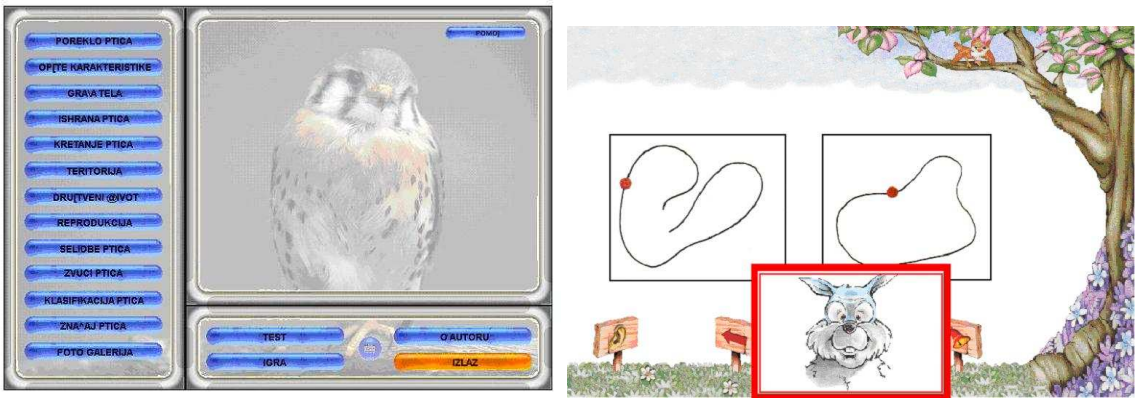
comparison to the control one. On the ground of these results we could make a conclusion that by implementing strategies for increasing efficiency in teaching Informatics, in which the use of computers is essential, students' motivation for work is considerably increased:



(a) Diagram of initial empirical distribution of knowledge level (b) Diagram of final empirical distribution of knowledge level

Figure 3:

Teachers or those participants in the process of e-system building who deal with designing, choose among different kinds which can be grouped in four types of basic models in organization of teaching contents.



(a) Intro-screen of educational software "Birds" for primary school [10] (b) Educational software "Math games" for Preshool [11]

Figure 4:

5 Summary

Educators (all those who participate in the system design) should be aware of some key factors in the teaching process while designing learning modules in order to put a system for electronic learning into practice:

- designing should be a cooperative process which incorporates graphic designers, language instructors etc.,
- a system should be flexible enough to allow students reading material and improving their understanding of the contents,
- it is necessary to develop courses' materials which enable giving recurrent information and group learning,
- it is necessary to use browser limitation cleverly, as well as hardware and software support in order to reduce methods of learning,
- advantages of external sources of information (e.g. hyper text links) should be accepted.

If teachers want to comprehend all needs and communication styles of all students they have to ask for continual recurrent information. New technologies and teaching methods have improved a traditional role of teachers in the learning process. However, teachers still have a great responsibility in stimulating students' interests, concerning motivation and selection of themes for learning in "Internet classrooms" [4].

In the near future the vision of global knowledge where the information which should be made available to public becomes "public goods" will come true. Such a tendency is called a "tendency of open approach" and it supports the idea of open electronic approach. Traditional learning is becoming absolutely insufficient without the implementation of modern electronic models for learning.

References

[1] Dragana Glušac, doctoral theses "Methodical and Didactical Issues of Efficiency in Teaching Information Technology", Zrenjanin, 2005.

[2] "What is Distance Education" by Virginia Steiner, DLRN 1995.

[3] http://wintel.fon.bg.ac.yu/ProgramIstrazivanja.htm

[4] Chris Shull, Noam Arzt, and Dan Updegrove:Blood, Sweat, and Tears on the Distributed Computing Trail, http://www.hln.com/noam/CEM9532.pdf

[5] http://www.poslovniforum.hr/about02

[6] Berlinska deklaracija o otvorenom pristupu znanstvenom znanju, http://eprints.rclis.org/archive/00000965/01/prijevod_berlinske_deklaracije.pdf

[7] UNICEF (2001), "Sveobuhvatna analiza sistema osnovnog obrazovanja u SRJ", Beograd.

[8] http://www.ekonomist.co.yu/magazin/ebit/

[9] "What is Distance Education" by Virginia Steiner, DLRN 1995.

[10] "Distance learning system model projecting", Mr Dragana Glušac and others, IEEE Mipro 2004.

[11] ww.icus.net/elearning/elearnstandards.shtm

[12] Slavomir Stankov, Ana Ban, "Pristupi i trendovi u standardizaciji E learning-a", Split 2004.

[13] Vesin Ivica, diplomski rad "Ptice", mentor Dragana Glušac, TFMP Zrenjanin 2006.

[14] Dragana Glušac, "Matematika kroz igru" softver za magistarsko istrazivanje, 2000.

Dragana Glušsc
Technical Faculty "Mihajlo Pupin" Zrenjanin
Yugoslavia
E-mail: gdragana@tf.zr.ac.yu

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 322-326

Data-Mining Techniques for Supporting Merging Decisions

Lucian Hâncu

Abstract: The Mergers and Acquisitions transactions have increased during the last years, as business entities face multiple threats caused by globalization and are unable to exploit the opportunities offered by the global market. These transactions come as a solution for consolidating the position of one entity on the local, national or global market, but usually surprise the competition, which does not have any prepared strategy for surviving the rise of a stronger competitor. In order to help the entities decide to which company should merge, or to be aware of the fact that a competitor could merge in the near future, we have built a technique for supporting merging decisions, based on the financial statements analysis and the Web usage logs extracted from our multi-server search application. The model suggests the merge with an entity which share the same activity code with the initial entity, or has a related activity code, according to the Business Dependency Map (derived from the Web search logs).
Keywords: Financial-statement analysis; Web usage mining; Merge decisions.

1 Introduction

The numerous examples of recently completed mergers (Arcelor and Mittal for the global steel industry [1], or Catex Calarasi and Serca [2] for the national textile industry) illustrate an increasing interest in the merging and acquisitions transactions proved by companies all over the world. The mergers or the acquisitions come as a solution for consolidating a market position (in the case the two companies share the same business activity) or to have access to new markets, when the two companies are business partners and have different activity domains.

The last decade's development of the Web technologies made available a large amount of valuable material that can be easily browsed and digested by humans, or indexed by search engines. The online availability of the Romanian companies' financial statements significantly eases the development of competitive analyzes, with the aim of improving the business community's knowledge of the competition and to predict its moves.

In this article, we consider the case of analyzing the Romanian entities' profitability of the capitals. We compute the profitability by automatically analyzing the online financial statements. We suggest merger decisions for companies whose profitability is below the average profitability on the sector. Furthermore, we make use of Web usage mining techniques [7] in order to derive dependencies between various sectors of the economy and suggest mergers between companies from related business activity sectors.

The paper is organized as follows: the subsequent section presents the method of analyzing the publicly available financial statements of the Romanian companies, whereas the third section discusses the method of building a map of dependencies between the various sectors of the economy. The fourth section highlights the results of classifying the entities according to the profitability of the capitals, the derived map of business dependencies and our technique of suggesting merging decisions. The last section points out the conclusions of our research and directions for future development of our methods.

2 Financial statements analysis

During our previous research, we have investigated various methods of collecting [5] and classifying [6] the financial statements of the Romanian entities. Official sources (like the Ministry of Finances and the Registry of Commerce) publish online financial information, which is retrievable by automatic filling of the Web search forms. Therefore, we can easily download large amounts of financial data for the purpose of further analyzes.

In this article, we analyze the Romanian entities' profitability of the capitals (gross profit divided by the total capitals), as the main indicator for predicting future merging operations. We aim to suggest the merger between two companies, either sharing the same activity code (according to the Romanian CAEN classification [3]) or having dependent activity codes (in the subsequent section, we shall present a technique for deriving these dependencies). The consequence of the merger would be the improvement of the resulting company's profitability of the capitals.

In order to accomplish our purpose, we compute each entity's profit margin and the average of this indicator on each one of the available CAEN activity codes. We consider that a business entity requires a merging if its

indicator is *far below* the average indicator of its activity group. In addition, the mentioned company should merge with an entity whose profitability of the capitals exceeds the average measure on the sector.

3 Dependencies between entities

A major disadvantage of the financial statements' analysis is that it does not consider the dependencies between the entities of the economy. Instead, it analyzes each single entity separately from its competitors and from its clients or suppliers. For improving our merger suggestions, we should also predict whether a sector of the economy is related to other sector, and propose mergers candidates from linked sectors. Hence, we need to compute the dependency map of the various sectors of the economy: we call that the entity A depends on the entity B, if either A is a client of B or A is a supplier for B).

These dependencies are crucial for suggesting mergers between companies, as the merging usually takes place between two competitor companies (in order to strengthen the position of the resulting company on the market), or between a producer and a customer (in order to have access to the customer's market and reduce costs). Although the list of the clients and the suppliers of a company is a private asset of that company and hard to be obtained from any third source, we can still predict the sectors of the economy in which the company acts as a supplier or as a client.

We shall illustrate the technique by an example: let us consider the case of an aluminum smelter and try to guess possible mergers for our business entity. The aluminium smelter gathers material from a *bauxite mine* and it delivers the aluminum to an *airplane producer*. Therefore, it could be interested in the merge with either a bauxite mine or an airplane producer, as a measure for reducing the costs of the acquired bauxite or the costs of producing the airplane and obtain larger profit margins. Based on our example, we shall conclude that the aluminum smelting activity (whose activity code is 2742, according to the CAEN Rev-1 classification [3]) depends on the bauxite mining activity (CAEN code: 1320) and the airplane building activity (CAEN code 3530) depends on the aluminum smelting activity.

Our method of predicting the dependencies between various sectors of the Romanian economy consists in analyzing the Web logs of our Business Information Search Engine Application d394.eu [4]. The Web logs contain a daily-based and server-based list of performed queries: a query consists in one or more financial codes of Romanian companies. Therefore, a dependency matrix can be easily computed by analyzing the logs and classifying the financial codes of the companies into their activity codes, according to [3]. In the second part of the next section, we present our results of analyzing the Web search logs and their usage in suggesting merger operations between the Romanian entities.

4 Results

4.1 Results from the analysis of the financial statements

The gathering of the publicly available financial statements of the Romanian entities resulted in the building of a database containing *518.409* active entities, which have regularly published their financial statements during the last 5 financial years. By analyzing the available information and computing the profitability of the capitals (gross profit divided by the total capitals), we have calculated the average indicator for each activity code (according to the CAEN classification) and outlined the best profitability indicator of each CAEN code.

The results, depicted in [Table 1], show the CAEN codes having an average profitability of the capitals above 1,200%. The table also enlightens the highest profitability of the capital for each activity code at the end of the 2006 financial year. The data impresses, as there is an important difference between the average profitability of the capital of the most part of activity codes (less than 100%) and the average profitability of the ones depicted in the table (greater than 1,200%). The analysis reveals that the Romanian economy has some entities with a very high indicator of profitability of the capitals (for instance, 847,732.90% - the profitability of an entity whose activity code is 6312 *"Deposits"*). The fact appears as there are several privately-owned companies who exhibit only a minimum required capital of *5 RON*.

Our study also shows that privately-owned businesses can provide a much higher profitability when compared to other investment alternatives (stock exchange or mutual funds). Even so, their lack of capitalization makes them vulnerable to hostile takeovers and decreases the corresponding investors' attractiveness.

CAEN Code	Average Profitability of Capitals (%)	Highest Profitability per CAEN Code (%)
0122	1533.93	239518.00
2411	1616.52	65595.83
2615	4366.41	125117.20
2861	2348.89	21896.67
4511	1895.94	395142.50
6010	1913.49	156028.50
6220	2701.85	98701.00
6312	4441.15	847732.90
6523	1574.28	109585.60
7320	1204.55	52688.89
7415	2287.31	158053.40
8022	2259.93	13088.93
9212	1351.68	64280.20

Table 1: The sectors with the highest profitability of capital (December 2006)

4.2 Analyzing Web Logs

We have analyzed our daily-based and server-based collection of Web search logs, gathered from our Multi-Server Search Application. The collection contains *267.713* log entries, both successfully and unsuccessfully searches (server errors or client mistakes). The map of Business Dependencies (see Figure 1) results as follows: for each two subsequent entries in a log file we consider that the activity code (according to the Romanian CAEN classification) of the second entry depends on the activity code of the first entry, therefore it would be probable that an entity from the first CAEN class would be either a client or a supplier of an entity from the second CAEN class.

The more intense is the color corresponding to the dependency between the two CAEN classes - the highest is the probability that the second class (depicted as columns of the Figure 1 map) depends on the first class (depicted as rows in the map). The computed business map shows an increased business activity corresponding to the region of the 4*** and 5*** CAEN groups (see the excerpts of the Figure 2).

To some extent, our automatically-generated map reflects the reality of the Romanian business economy: the 4*** group (which is the fifth group from left to right and from top to bottom, according to the margins of Figure 1) denotes utility supplier entities (electricity, gas), whereas the 5*** group groups the entities from the commerce sector. It is straightforward that almost every business sector depends on the entities of the 4*** CAEN class; it is also highly expected that there is an increased dependence between the entities from the 5*** CAEN class (depicted as the red cells in the above enumerated figures).

An interesting result is that there are no dependencies between some regions of the maps: we explain this finding as the entities of some business activity codes do not have intense business activity with other sectors of the economy. For instance, the 0*** CAEN class (agricultural sector) and the 9*** CAEN class (the last one depicted in the Figure 1 - from left to right and from top to bottom) are scarcely represented on the map. We should also take in consideration the fact that we have built the model based on Web log entries, which means that some entities of the economy could be easily left outside the log entries, as users did non search for information on those entities.

4.3 Suggesting mergers

Once we collect the financial information, analyze it and generate the business map for supporting the mergers, we simply find an entity with which our *A* company should merge. We point out that the purpose of the merge should be the increase of the profitability of the capitals of the first company (we call it *company A*). For that company, we compute the list of the candidate merger entities: the list contains the entity which has the highest profitability of the capitals in the same activity code as our A company and the companies with the highest profitability of the capitals in the sectors which come in a dependency relation with the sector of the *company A*, according to the Business Dependency Map shown in Figure 1. The company which exhibits the highest profitability of the capitals from the candidate companies becomes the suggestion for the merger transaction.

The fact that the candidate companies are top performers in their sectors of activity (in terms of profitability of capitals) assures the improvement of the indicator on the resulting company, as compared with the initial company

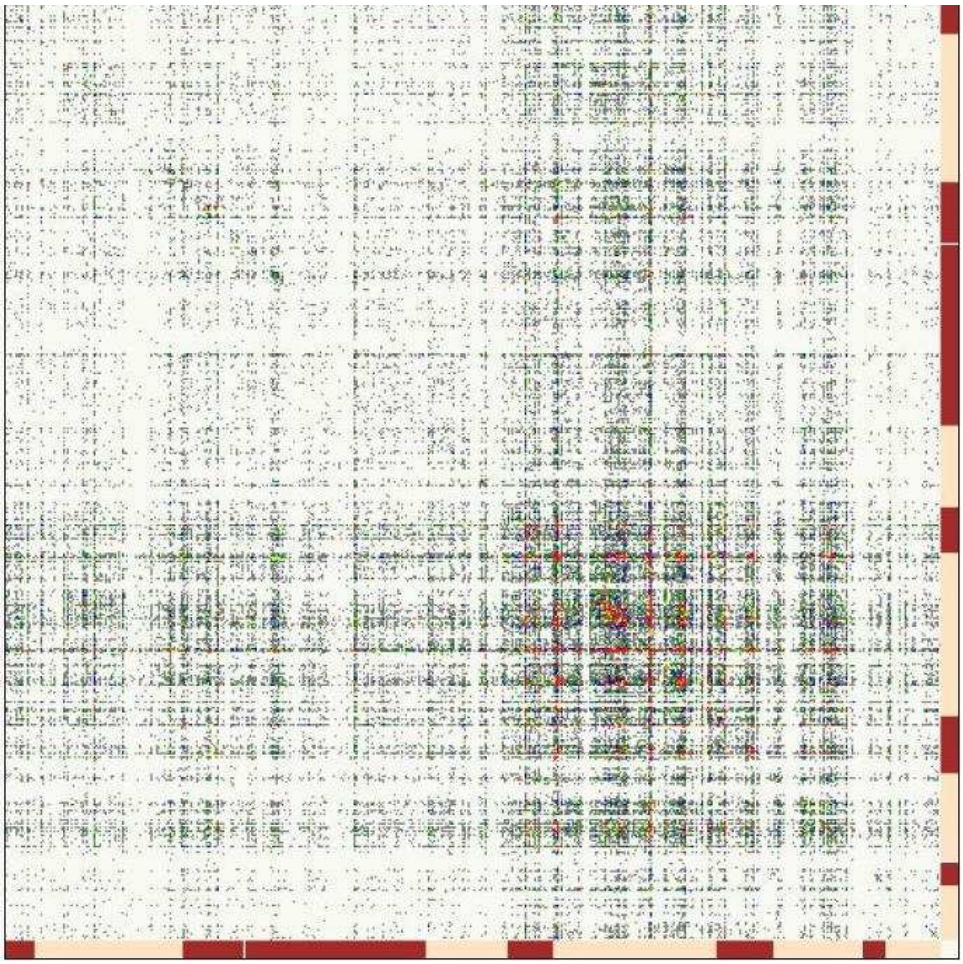


Figure 1: The Map of Business Dependencies

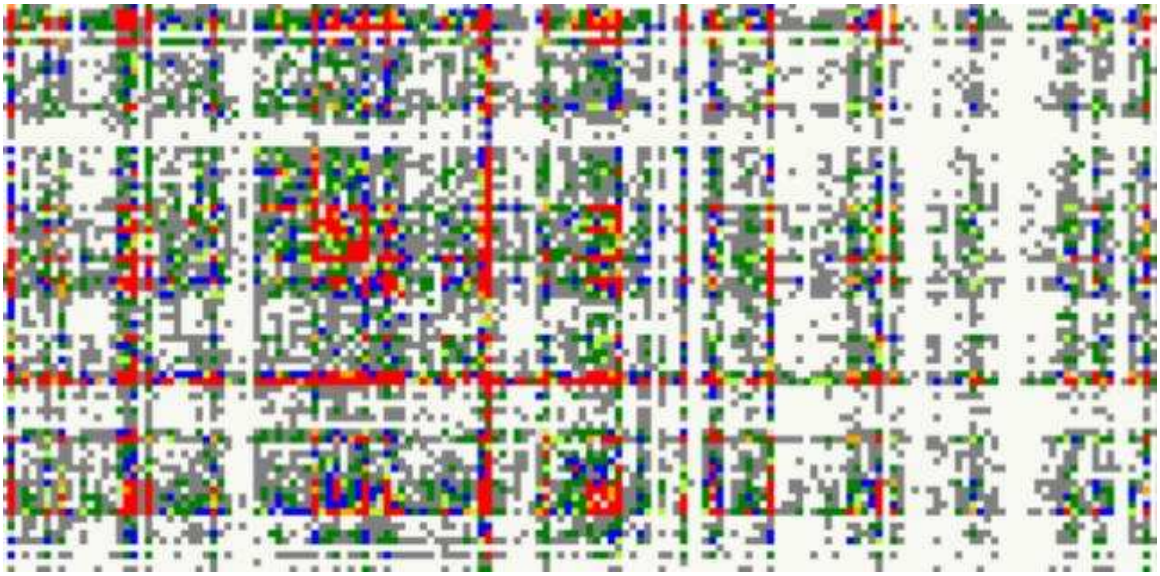


Figure 2: Map of Business Dependencies: Excerpts from Figure 1

A. In terms of profitability of the capitals, the suggested merger candidate (we shall name it *B*) will rather not be interested in a merge operation. The merge could be an advantage for both companies if the resulting company's position in the market is significantly strengthened, when compared to the individual companies before the merge.

5 Summary and Conclusions

We presented a technique for suggesting merging operations based on the financial statements of the Romanian companies and the Web logs collected from our Business Information Server Search Application D394. We chose the profitability of the capitals as the definite criterion for suggesting which company should merge with which company. The indicator was computed using the financial statements available at the end of 2006.

The model can be easily extended to a multi-year analysis of the financial statements. We also plan to use other indicators (like the position of the companies on the market, or the total intangible assets of each company) to improve the merger suggestion model. The use of the total intangible assets enhances the model to suggest also possible acquisitions or absorptions on the market. A model of predicting merger or acquisition operations will definitely be useful for the business community, as a merger announcement usually comes as a surprise in the market, with little chance of reaction for the competing entities.

References

[1] R. Miller, "Global Steel is Coming Together", *Business Week*, September 13, 2006.

[2] G. Sarcinschi, "O firma a 'regelui confectiilor' absorbita de Catex Calarasi", *Business Standard*, December 2, 2007.

[3] National Institute of Statistics, "The classification of the National Activities - CAEN Rev-1, *http://www.ins.ro*, 2002.

[4] SoftProEuro, "Declaratia 394", *http://www.d394.eu*.

[5] L. Hancu, "Enhancing the Invisible Web", *KEPT 2007 International Conference*, Cluj-Napoca, Romania, June 2007.

[6] L. Hancu, "The Pre-Accession Competitiveness of the Romanian Software Companies", *International Conference Competitiveness and European Integration*, Babes Bolyai University, Cluj Napoca, Romania, October 2007.

[7] M. Konchady, "Text Mining Application Programming - Programming Series", *Charles River Media*, May 2006, pp. 197-202.

Lucian Hâncu
SoftProEuro s.r.l.
Cluj-Napoca
E-mail: lhancu@softproeuro.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 327-332

Modelling of the Distributed Databases. A Viewpoint Mechanism of the MVDB Model's Methodology

Daniel I. Hunyadi, Mircea A. Musan

Abstract: Over the past years, most of the research dealing with the object multiple representation and evolution has proposed to enrich the monolithic vision of the classical object approach in which an object belongs to one hierarchy class. In databases, much work has been done towards extending models with advanced tools such as view technology, schema evolution support, multiple classification, role modelling and viewpoints. In particular, the integration of the viewpoint mechanism to the conventional object-oriented data model gives it flexibility and allows one to improve the modeling power of objects. The viewpoint paradigm refers to the multiple description, the distribution, and the evolution of object. Also, it can be an undeniable contribution for a distributed design of complex databases. The motivation of this paper is to define an object data model integrating viewpoints in databases and to present a federated database architecture integrating multiple viewpoint sources following a local-as-extended-view data integration approach.
Keywords: object-oriented data model, OQL language, viewpoint schema, LAEV data integration approach, MVDB model, federated databases, Local-As-View Strategy.

1 Introduction

Object-oriented databases are becoming more and more popular for applications to support the complexity and the irregularity of the real-world entities. Moreover, with the expansion of the distributed technology and the Internet, new needs related to data sharing and data exchange appear. Thus, the development of advanced database models is required. Object-oriented technology seems to be the keystone of this evolution. Hence, much work has been done recently towards extending object-oriented database models with advanced tools such as view technology, schema evolution support, multiple classification, role modelling and the viewpoint paradigm. All these extensions require more flexible and powerful constructs than are currently supported by existing object -oriented models [10].

The viewpoint paradigm is an active subject of research in many areas such as software engineering [1], knowledge representation [2], database systems [3], [4], web applications [5], etc. In DataBases (DBs), we notice few works on the integration of the viewpoint concept into the data models. Most of these works consider the view and the role mechanisms. Views [6], [7] are external schemas that provide the user with a part of the global schema, a kind of viewpoint on the description of its entities. Roles [8], [3], [9] deal with the multiple aspects that an object acquires and loses during its life-time within a unique representation. In the context of our work, viewpoints offer several descriptions to the same Universe of Discourse (UoD). Each description is not a view, but a partial representation of data according to a given point of view. The various partial descriptions are supported by database schemas that together provide the global schema of the same real world data. Objects can be described according to one or more descriptions, as a kind of role within a multiple data representation. Achieving such an approach requires a distributed environment and, more precisely, a federated database system that permits the integration and the collaboration of a collection of databases.

In this paper, we report an ongoing research we are engaged in [6]. Our work is aimed at extending object-oriented database technology to accommodate multiple and distributed modelling of data. The paper is structured as follows. Section 2 provides an overview of the viewpoint approach used in the several fields of computer science. A comparison of the integration of the viewpoint paradigm in database modelling is given in Section 3. In Section 4 and Section 5 we present the methodology and formalization of the MVDB (Multi-Viewpoint DataBase) model, respectively. The proposed model is an extension of the conventional object data model with the viewpoint mechanism. It allows developing a schema as a multiple description of an UoD. This description consists of translating several abstractions of this universe, using a basic formalism for the multiple data descriptions. Section 6 presents the consistency and objects evolution in the MVDB model and we give the general architecture of a federated database system, called MVDB system, that uses an adapted LAV approach to integrate viewpoint sources. Section 7 concludes our work.

2 The Viewpoint Approach

In computer science, most of data modeling systems don't deal with the variety of perceptions related to the same UoD and develop tools to create a single model for a single vision of the observed world. The viewpoint approach is opposed to this monolithic approach and makes it possible to model the same reality according to different points of view.

The viewpoint approach is constructed on the conjunction actor/information. Therefore, it is necessary to include the actor in the action. We thus define a viewpoint as "a conceptual manner binding, on the one hand an actor who observes and, on the other hand, a phenomenon (or a world) which is observed". Many actors can observe the same UoD and produce various viewpoints on it. These last can be considered in several manners illustrated in Figure 1.

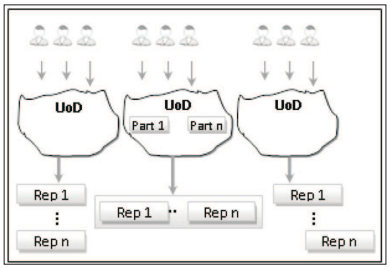


Figure 1: The different manners to consider viewpoints on an UoD

Uniform viewpoints: in this case, all the actors have the same vision of the UoD and produce equivalent representations. For example, let us consider many research teams, each one uses a different data model and considers it as the best one to represent a project.

Complementary viewpoints: in this case, each actor sees a part of the UoD and provides a viewpoint on it. Each viewpoint is a partial and coherent representation. The various representations which rise from the various actors are complementary and their union is a complete and coherent representation of the UoD.

Comparable viewpoints: in this case, the actors produce comparable representations according to the generalisation/specialization meaning. Within the framework of our study we are interested in the second interpretation of the relation "actor-world", which supposes that the various viewpoints on the same UoD are partial but complementary representations of it.

The viewpoint mecahnim has been integrated into various contexts and used to solve different problems. Most works in the literature dealing with the viewpoint notion in object-oriented and conceptual modelling are much more pragmatic. In the following, we identify the main objectives in integrating viewpoints into computer systems. Note that there is no single use of this concept that includes all of these objectives.

• The viewpoint as a means of providing multiple descriptions of an entity: the viewpoint concept seems to naturally result from the multiple views of objects of a specific study. As a matter of fact, a real world entity can have many behavioral contexts and many states from which the notion of multiple descriptions has been derived. Recently, the viewpoint paradigm has also been applied to web data in representing and viewing multidimensional information; that is information that may assume different facets under different contexts [5].

• The viewpoint as an approach for the modelling and distributed development of systems: many authors state that the modelling of complex systems as defined in cannot be handled with the same techniques as used for simple systems. However, the modelling of a complex system cannot be a centralized task based on a single formalism. Solutions based on logical systems are generally used to permit this correlation.

3 Related Works

In the field of databases, the concept of viewpoints is mainly investigated within the concept of views and roles in the object-oriented database community. Most of the research works propose enriching the monolithic vision of the traditional object-oriented approach in which an object belongs to one and only one hierarchy class. They deal with the objects evolution and with the existence of multiple views of the same data. In this section, we briefly examine some proposals which present roles and views, and then we present an overview of our viewpoint approach.

3.1 Views

Various view models have been proposed such as the multi-view model of [10] and the view model of [1] and of [7]. In these works, views are exploited to allow different applications to see the same database according to different viewpoints. The viewpoint concept here supports external schema, which is the third level of the ANSI architecture standard upon which the construction and the use of relational database systems and the later object-oriented ones are centred. Many problems arise, such as how a view schema (view class) is inserted in a global schema (class hierarchy) and whether an instance of a view owns an identity. A view can be treated as a database, but it does not preserve an object identity. Rundensteiner and Bertino [7] introduce the concepts of the multiview and schema view, respectively. These provide the capacity to restructure a database schema so that it meets the need of specific applications. They present support for view design by automating some tasks of the view specification process and by supporting automatic tools for enforcing the consistency of a view schema. Indeed, different views of the same object are allowed, depending on the context in which the object is considered. Here views preserve an object's identity, but the different instances of the same object are independent. All these models consider the viewpoint as a view defined with the aim of adapting an existing structure to new needs.

3.2 Roles

Objects with roles have increasingly been studied by several authors [8], [3], [10]. Roles are useful for supporting objects with multiple interfaces that can be dynamically extended to model entities which change their behaviour, and the class they belong to over time. This task presents many problems such as uniqueness of objects identifier, strong typing, persistence, late binding, etc. in response to the role handling problem, several approaches have been introduced. In particular, the intersection-class-based and the role-hierarchy-based approaches are the most popular. The first approach simulates the objects multiple classification and dynamic restructuring by creating an intersection class to reflect the structure of a multiply-classified object. A separate class must thus be defined for every combination of roles. This simulation adheres to the constant that an object belongs to exactly one class at a time. This can present many problems: the class hierarchy may grow exponentially and the dynamic object classification is a tedious task. The role hierarchy-based approach, however, has been adopted in many extended object-oriented database systems [10]. A role hierarchy is a tree of special types called role types. The root of this tree defines the time-invariant properties of an object. The other nodes represent types (roles) that an object can acquire and lose during its lifetime. The notion of roles is thus essential to support object extension, but is also useful to model situations where one real world entity may exhibit different behaviour in different contexts without changing its identity within a unique representation. Objects can therefore have several contexts, i.e. a kind of viewpoint that it acquires and loses dynamically.

4 The Methodology of the MVDB Model

The MVDB is an object-oriented data model with an extension by concepts and mechanisms which allow the multiple, evolutionary and distributed representation of a database schema. This representation confers to an UoD several partial and complementary representations. Each partial description is based on a first description of the entities and extends it according to a given viewpoint. The multiple and evolutionary representation overcomes the restriction of the single and fixed object instantiation link. The distributed representation fulfills the requirements of the current applications of the distributed and decentralized development of databases.

We adopted the object-oriented model as the common model for the various database schemas. This choice is justified by three principal motivations. First, the application of object-oriented concepts in system architectures provides a natural model for autonomous and distributed systems. Second, the object technology has been used in multidatabase systems to a finer level of granularity. Third, the expression and struc-turing power of the object-oriented approach goes with the objects modelling features in the MVDB model, such as the multi-instantiation mechanism that permits an object to have more than one instance.

The methodology of the MVDB model relies on the following ideas:

• the viewpoint concept is considered as an inherent concept of the data model and not as an augmented mechanism on it;

• a database schema is a multiple description of the same UoD according to various viewpoints. A database schema is thus viewed as a set of VP schemas, as shown in Figure 2. Each VP schema represents an aspect of the data description and is held by an independent database system;

• the VP schemas construction is based on a basic one called the referential schema. This last holds basic data on the real word entities shared by all the VP schemas;

• objects in the referential base are global. Global objects have a basic description in the referential base and one or more descriptions according to viewpoints;

• objects evolution is held by allowing entities to acquire or lose partial descriptions in the different viewpoint schemas while preserving their identities.

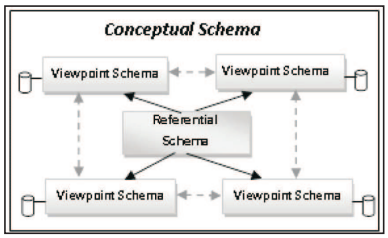


Figure 2: The viewpoint approach

We point out that object identity is a central notion in our approach. It is the same object described in many ways according of its membership in the various VP schemas. However, in order to ensure the components autonomy, local objects can be created and managed locally by VP databases. Local objects are objects with a single description according to one viewpoint and can't be accessed at the global level. VP databases are complementary and provide a global distributed database called multi-viewpoint database. A coherent exploitation of this global database is then recommended. Generally, these features are particularly needed in large complex applications of the industrial world. As a matter of fact, companies are logically distributed into offices, departments, working groups, etc. Consequently we can deduce that the data are also already distributed. Each unit in the company must manage the relevant data for its operation and should be able, if necessary, to reach remote data that exist in the other units. The data in the various units are complementary and operated upon by collaborating users.

We illustrate the viewpoint approach through a simple modelling example. It concerns the representation of a laboratory's scientific staff (see Figure 3). This is composed of a referential schema and two viewpoint ones. The referential schema consists of the common information shared by all the viewpoints. We are particularly interested in the teaching and research activities of each member of the laboratory. Let us consider the Research VP and

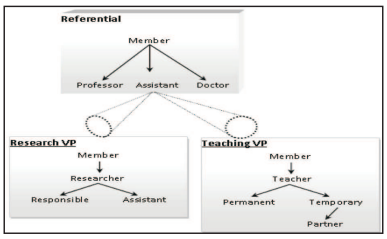


Figure 3: A multi-viewpoint modelling example.

the Teaching VP. Each viewpoint is an object-oriented schema that contains only information that is relevant to it. The Research VP, for example, is a hierarchical description of the laboratory's members according to their research activity. Each member can have, simultaneously, a basic description at the referential level and one or two viewpoint descriptions according to his/her teaching and research activities. For example, an one member is presented with oid "E1" in Figure 4, is a professor, permanent teacher and responsible of research topics. E11 and E12 are his identifiers at the VP schemas.

5 The MVDB Architecture

We have noticed above that the viewpoint approach to databases requires a distributed environment. Distributed systems [9], [5] have become increasingly important because of requests for organization and the growth of advanced techniques in the network management. These systems are characterized by three orthogonal dimensions: distribution, heterogeneity and autonomy. In this paper, we do not deal with the heterogeneity dimension.

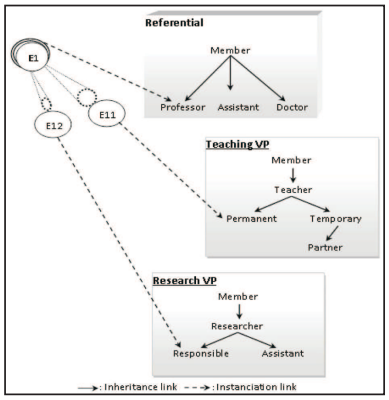


Figure 4: The multi-viewpoint object representation.

According to the autonomy dimension, [9] propose a classification most commonly applied to the distributed systems. These are divided into two families: non federated or tightly-coupled database systems and federated or loosely-coupled database systems. In tightly-coupled database systems all the various database schemas are integrated in only one global schema. The integration of the components makes these latter lose all their autonomy. Indeed, there is only one management level where all the operations are carried out in a uniform way. Then no distinction is made between the local and the global use of data. Thus, this approach does not meet the viewpoints structuring needs. As a matter of fact, a federated system consists of the integration of many autonomous and interdependent database systems. Thus, in contrast to the previous approach, a federated database does not support a global schema. Its main objective is to ensure the autonomy of the component databases and to privilege their management and their independent handling. The federation is an appropriate architecture to support the viewpoint approach. However, what about the data integration strategy that will be used?

In a federated system two strategies are used to integrate independent databases in a unified logical global schema: the Global-As-View (GAV) strategy that defines the global schema as a view over the local schemas and the Local-As-View (LAV) strategy that defines the local schemasas views over the global schema [9]. We are particularly interested in the LAV architecture that will be adapted to our architecture.

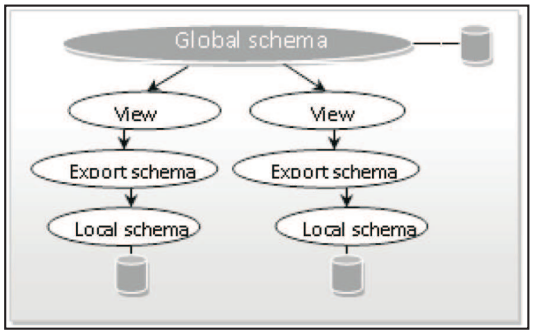


Figure 5: The LAV data integration approach.

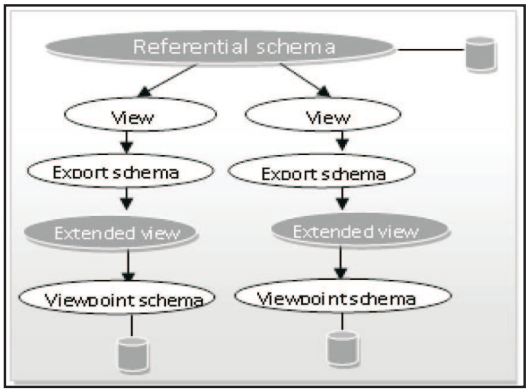


Figure 6: The LAEV data integration approach.

The Local-As-View (LAV) strategy, presented in Figure 5, consists of defining the local sources as views over the global schema. This presents two principle advantages: a local change to a data source is easily handled and the heterogeneity of the different components is supported. The LAV process is more adaptable to the data model we have defined above. However, in our case, local schema called viewpoint schema is an extended view over the global schema called the referential schema. We recall that a viewpoint schema is a partial description of data according to a viewpoint. A Local-As- Extended-View (LAEV) process is then used in our system (see Figure 6).

6 Conclusion

In this paper, we have proposed a structural object database model that integrates the viewpoint paradigm. This approach refers to the evolution, multiple description and distribution of objects. Also, it can make an undeniable contribution for the distributed design of complex databases. However, the same UoD can be described in a distributed fashion by different database schemas. Each one of these presents the entities according to a single viewpoint. A federated environment instead of a centralized one has been chosen to achieve our approach. Future work would concern the development of a data definition and manipulation language for the MVDB model, which is an extension of the OQL language. In addition, it would be interesting to develop an expression language to specify integrity constraints at the federation level.

References

[1] P. J. Charrel, D. Galaretta, C. Hanachi, B.Rothenburger, *Multiple Viewpoints for the Development of Complex Software. Proceedings of the IEEE Int′l Conference on Systems, Man and Cybernetics*. (1993), pp. 556-561. Le Touquet, France.

[2] O. A. Bukhres, A. K. Elmagarmid,*Object-Oriented Multidatabase Systems*. Prentice-Hall, (1996), Englewood Cliffs, NJ.

[3] S. Coulondre, T. Libourel, *An Integrated Object-Role Oriented Database Model*. Data & Knowledge Engineering 42(1), (2002), pp. 113-141.

[4] H. Naja, Cedre: *un modéle pour une représentation multi-points de vue dans les bases d′objets*. Doctoral thesis, University of Henri Poincaré, Nancy 1, (1997).

[5] M. Gergatsoulis, Y. Stavrakas, D. Karteris, A. Mouzaki, D. Sterpis, *A Web-Based System for Handling Multidimentional Information through MXML*. Lecture Notes In Computer Science (LNCS 2151), (2001), pp. 352-365, Springer-Verlag.

[6] S. Abiteboul, A. Bonner, *Objects and views*, Proceedings of the Int′l Conference on Management of Data, ACM SIGMOD, (1991), pp. 238-247. Denver, Colorado.

[7] E. Bertino, *A View Mechanism for Object-Oriented Databases.* Proceedings of the 3rd Int′l Conference on EDTB'92, (1992), pp. 136-151.

[8] A. Albano, R. Bergamini, R. Ghelli, R. Orsini, *An Object Data Model with Roles*. Proceedings of the Int′l Conference on Very Large Database, (1993), pp. 39-51. Dublin, Ireland.

[9] Fouzia Benchikha, Mahmoud Boufaida, *The Viewpoint Mechanism for Object-oriented Databases Modelling, Distribution and Evolution*; Journal of Computing and Information Technology - CIT 15, 2007, 2, 95-110, doi:10.2498/cit.1000692

[10] G. Gottlob, M. Schrefl, B. Rock, *Extending Object-Oriented Systems with Roles*. ACM Transactions on Information Systems 14(3), (1996), pp. 268-296.

Daniel I.Hunyadi, Mircea M. Musan
Lucian Blaga University
Department of Computer Science
5-7 Ioan Ratiu Str, Sibiu, Romania
E-mail: {daniel.hunyadi,mircea.musan}@ulbsibiu.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 333-336

The Influence of Parameters on the Phaseportrait in the Mixing Model

Adela Ionescu, Mihai Costescu

Abstract: The problems of flow kinematics are far from complete solving. Recently, the mixing theory issued in this field, and the mathematical methods and techniques developed the significant relation between turbulence and chaos.
In the previous works, the study of the 3D non-periodic mixing models exhibited a quite complicated behavior. In agreement with experiments, they involved some significant events - the so-called rare events. The variation of parameters had a great influence on the length and surface deformations.
The 2D cases, both periodic and non-periodic, are simpler, but significant events can also issue for irrational values of the length and surface versors, as is the situation in 3D case. In the graphic analysis recently realized, in 2D case, the mixing has also a nonlinear behavior and the rare events can appear.
This paper is continuing the computational analysis for 2D mixing model, in a perturbed version. For the simulations there are used specific procedures and functions of Maple11. The conclusions will be further used for analyzing the mixing efficiency.
Keywords: turbulent mixing, stretching, folding, efficiency, phaseportrait

1 Introduction

The turbulence is an important feature of dynamic systems with few freedom degrees, the so-called "far from equilibrium systems", which are widespread between the models of excitable media. In this area two important theories are distinguished: the transition theory from smooth laminar flows to chaotic flows, characteristic to turbulence, on one hand, and statistic studies of the complete turbulent systems, on the other hand

The statistical idea of flow is represented by the map:

(1) $x = \Phi_t(X),\ with\ X = \Phi_t(t=0)(X)$

In the continuum mechanics the relation (1) is named *flow*, and it is a diffeomorphism of class C^k. Moreover, (1) must satisfy the relation:

(2) $0 \langle J \langle \infty, J = \det\left(\frac{\partial x_i}{\partial X_j}\right)$, or $J = det(D\Phi_t(X))$,

where D denotes the derivation with respect to the reference configuration, in this case **X**. The relation (2) implies two particles, X_1 andX_2, which occupy the same position x at a moment. Non-topological behavior (like break up, for example) *is not allowed*.

With respect to **X** there is defined the basic measure of deformation, the deformation gradient, **F**, namely [5]:

(3) $F = (\nabla_X \Phi_t(\mathbf{X}))^T,\ F_{ij} = \left(\frac{\partial x_i}{\partial X_j}\right)$,

where ∇_X denotes differentiation with respect to X. According to (3), **F** is non singular. The basic measure for the deformation with respect to **x** is the *velocity gradient.*

After defining the basic deformation of a material filament and the corresponding relation for the area of an infinitesimal material surface, we can define the basic deformation measures: the *length deformation* λ and *surface deformation* η, with the relations [5]:

(4) $\lambda = (C : MM)^{\frac{1}{2}},\ \eta = (\det F) \cdot \left(C^{-1} : NN\right)^{\frac{1}{2}}$,

with $C(= F^T \cdot F)$ the Cauchy-Green deformation tensor, and the vectors M, N are the orientation versors in length and surface respectively. The scalar form for (4), used in practice, is:

(5) $\lambda^2 = C_{ij} \cdot M_i \cdot N_j,\ \eta^2 = (\det F) \cdot \left(C_{ij}^{-1} \cdot N_i \cdot N_j\right),\ with\ \sum M_i^2 = 1, \sum N_j^2 = 1.$

The last condition is the condition for the versors.

The deformation tensor **F** and the associated tensors **C**, $\mathbf{C}^{-1}$ represent the basic quantities in the deformation analysis for the infinitesimal elements.

2 The perturbed model. Comparative analysis

Studying a mixing for a flow implies the analysis of successive *stretching* and *folding* phenomena for its particles, with the influence of parameters and initial conditions [5].

In 3D non-periodic case, the models exhibited a quite complicated behavior [1]. In agreement with experiments, they involved some significant events - the so-called "*rare events*", which correspond to the breakup of the simulation program. The variation of parameters had a great influence on the length and surface deformations [6].

The 2D case is simpler, but significant events can issue for irrational values of the length and surface versors. In [2] it was studied the periodic case, and in [3,4] a comparative analysis with non-periodic case was started. For the modified model

$(6)\dot{x_1} = Gx_2 + x_1,$

$\dot{x_2} = KGx_1 + x, \ -1\langle K\langle 1, 0\langle G\langle 1$

the solution of the associated Cauchy problem was found as some complex combination of exponentials where the parameter KG^2 is involved, therefore the length deformation analysis produced nonlinear phenomena.

In what follows, a modified perturbation of the classic 2D model is taken into account, in order to compare the behavior with the initial case. Namely, it is considered the following model:

$(7)\dot{x_1} = Gx_2 + x_1,$

$\dot{x_2} = KGx_1 + G\cdot(x_2 - x_1), \ -1\langle K\langle 1, 0\langle G\langle 1$

and it is searched a comparison from the analytical standpoint with the initial model [5]:

$\dot{x_1} = Gx_2,$

$\dot{x_2} = KGx_1, \ -1\langle K\langle 1, 0\langle G\langle 1$

There are used some specific procedures and functions of the soft MAPLE11, in discrete time, for simulating the behavior of the trajectories-solutions. As the model is non-autonomous, the procedure "*phaseportrait*" was replaced with "*DEplot*" (or *dfieldplot*). The procedures are based on the solver of *rkf45* type (Runge-Kutta method of fourth/fifth order).

There were taken into account three values for the parameters G and KG, namely:

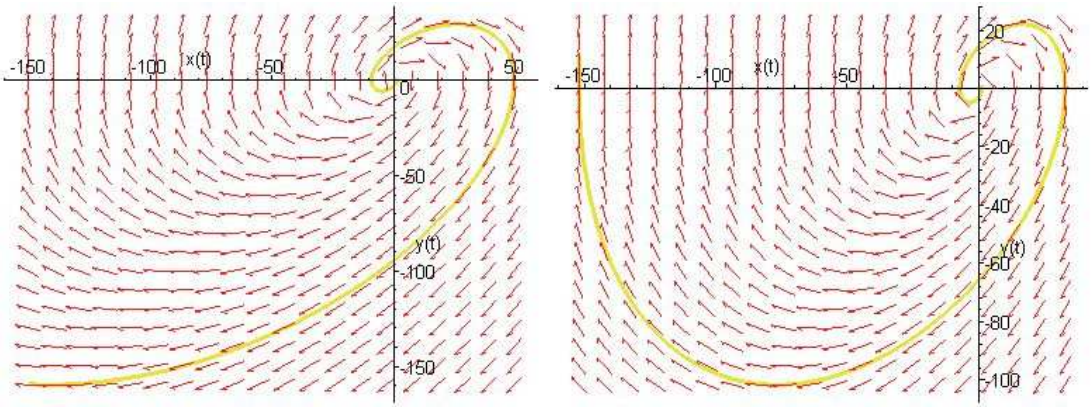
$a)G = 0.25, KG = -0.035;$

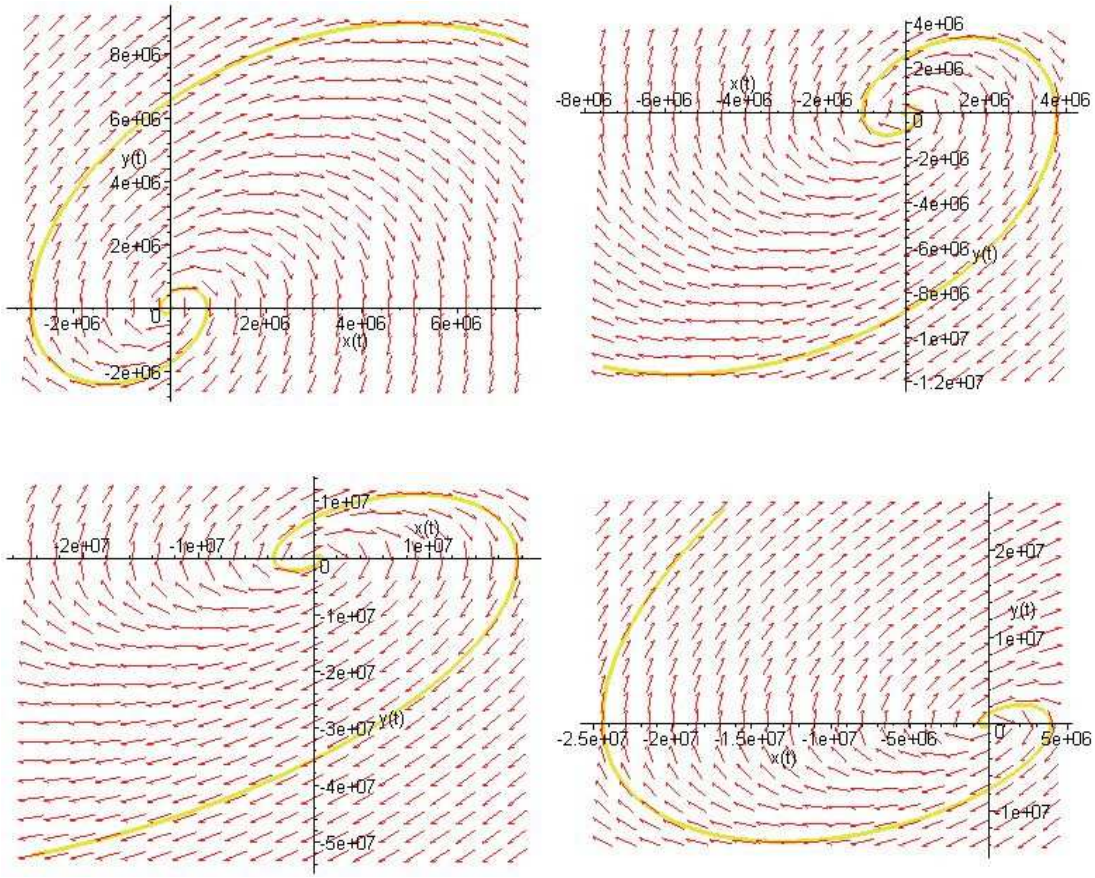
$b)G = 0.755, KG = -0.65;$

$c)G = 0.85, KG = -0.25.$

For each case, a little modification of KG was taken into account, for pointing out the comparison, namely: $a.1)\ KG = -0.085, b.1)KG = -0.85$ and $c.1)KG = -0.05$ respectively. For the initial case there were not noticed special evolutions. The figures numbered with two digits represent the corresponding case of the modification of KG in each of the cases a)-c).

For the perturbed case, there are exhibited both of cases for KG, and the simulations were as follows:





3 Remarks and Conclusions

1. It must be noticed how the phase portrait changes when the model is perturbed. In the initial case, the origin, (which is an equilibrium point), is *center*, and in the perturbed case (7), it tends to became an *unstable focus*. The time scale is the same, 40 units. There can be small units, as well as large units, in agreement with the experiments.

2. For the perturbed model, it is noticed, as in [4], a nonlinear/ negative behavior. Also, there are involved large values in the behavior of the trajectories, that means the model can become *far from equilibrium*.

3. Getting over from the initial to the perturbed model and studying the evolution of the phaseportrait, it becomes necessary a *spectral analysis* of the mixing model, for getting a new standpoint on the study and studying the influence of parameters. This is a next aim.

4. Also, as next aim is to test other parameter values, and possible, to enlarge the time scale. In the perturbed case it is better observed how smooth is the influence, since a little perturbation, as in (7), involves important changes in the phase portrait.

References

[1] A. Ionescu, "The structural stability of biological oscillators. Analytical contributions", *Ph.D. Thesis*. Politechnic University of Bucarest, 2002.

[2] A. Ionescu, "Computational aspects in excitable media. The case of vortex phenomena", *Proceedings of the International Conference ICCCC2006* , Univ. of Oradea, pag 280-284, 2006

[3] A. Ionescu, M. Costescu, D. Coman, "A computational approach of turbulent mixing. The trajectories analysis", *Recent Journal,* University of Brasov, vol 8(2007), nr 3a(21a,b), pag 498-501

[4] A. Ionescu, M. Costescu, "Some qualitative features of 2D perturbed mixing model.", *Acta Universitatis Apulensis,* University of Alba-Iulia, no 15(2008), pag. 387-396.

[5] J.M. Ottino, "The kinematics of mixing: stretching, chaos and transport", *Cambridge University Press.,* 1989

[6] S.N. Savulescu, "Applications of multiple flows in a vortex tube closed at one end", *Internal Reports, CCTE, IEA.* Bucarest. (1996-1998)

Adela Ionescu*, Mihai Costescu*

Adela Ionescu, Mihai Costescu
University of Craiova, Romania
E-mail: adela0404@yahoo.com, miracos2003@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 337-342

Intelligent Autopilot Control Design for a 2-DOF Helicopter Model

Saeed Jafarzadeh, Rooholah Mirheidari, Mohammad Reza Jahed Motlagh, Mojtaba Barkhordari

Abstract: In this paper, we introduce a new intelligent control approach called Brain Emotional Learning Based Intelligent Controller (BELBIC). BELBIC is a controller based on emotional model of human brain which has been introduced not for so long. This controller has been applied to a nonlinear model of a helicopter. Feedback linearization method has also been applied to the system, and the performance of two controllers has been compared as an intelligent and a classical control method. An Input to State linearization method with some changes has been used to control the system. The performance of the controllers has been justified by the simulation.
Keywords: BELBIC, Reward, Feedback Linearization, Input to State Linearization.

1 Introduction

Most of the existing results for helicopter control have been based on the linearization model or through several linearization techniques [1, 2]. The linearized models of the helicopter have some unmodeled dynamics that make the proposed controls unreliable when applied to original nonlinear models. Using nonlinear techniques gives better response and improves the performance of controlled system.

A very thorough survey of linear techniques for helicopter control has been given by Garrad and Low [1]. Miniature helicopter control problems have also been discussed by Furuta *et al*. [3] and Kientz *et al*. [2]. The work of Pallett *et al*. [4] has served as the basis for our understanding of the helicopter model. Some other nonlinear control methods have been applied to the model which has been considered in this paper. Sliding mode Control [5] and robust control method [6] are some examples of these works.

In this paper we apply two methods to the system and compare their performance. First method is a new intelligent approach called Brain Emotional Learning Based Intelligent Controller (BELBIC). This technique is based on physiological structure of brain in excited situations and can be used in control engineering problems. Recently, there is rising tend to intelligent controllers and BELBIC is not an exception [7, 8, 9]. This controller has a certain structure, but it can be changed to achieve the control objectives. There are both continuous and discrete time BELBIC controllers [9, 10]. We use the continuous one to compare with another continuous time control method (feedback linearization). To design the BEBIC controller, we should choose the appropriate reward function according to physical aspects of the control problem, and tune the training coefficients of the controller to achieve desired control objectives. Second method is feedback linearization which is a classical control method. An input to state linearization has been applied to the system for this method. There has been applied a feedback linearization method to a model of a helicopter [11], but it is different from the model which we use here. Therefore, we design a controller by feedback linearization method for the helicopter system.

There is a complicated fifth order nonlinear state space model for the system. The system has two inputs and two outputs. The control objective is to track the desired set points of the outputs. The sections of the paper are as follows. Section 2 describes the fifth order model of the helicopter. Section 3 contains the theoretical results in BELBIC controller design and its tuning. Theoretical aspect and controller design for feedback linearization method has been brought in the section 4. Section 5 shows the simulation results and finally section 6 is conclusions.

2 Helicopter Model

Consider the nonlinear helicopter model as follows [6]:

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= a_0 + a_1 x_2 + a_2 x_2^2 + \left(a_3 + a_4 x_4 - \sqrt{a_5 + a_6 x_4}\right) x_3^2 \\
\dot{x}_3 &= a_7 + a_8 x_3 + (a_9 \sin x_4 + a_{10}) x_3^2 + u_1 \\
\dot{x}_4 &= x_5 \\
\dot{x}_5 &= a_{11} + a_{12} x_4 + a_{13} x_3^2 \sin x_4 + a_{14} x_5 + u_2
\end{aligned}$$

$x = [h\ \dot{h}\ \omega\ \theta\ \dot{\theta}]^T$ and $u = [u_1\ u_2]^T = [k_1 u_{th} - k_2 u_\theta]^T$

In these equations h is the height of helicopter above ground and is measured in meter. ω is the rotational speed of the rotor blades and is measured in radian per second. θ is the collective pitch angle of rotor blades and is measured in radian. u_{th} and u_θ are the input to the throttle and the input to collective servomechanisms respectively.

3 Feedback Linearization

Feedback linearization has been used successfully to address some practical control problems. These include the control of helicopter, high performance aircraft, industrial robots, and biomedical devices [12]. The central idea of the approach is to algebraically transform a nonlinear system dynamics into a (fully or partly) linear one, so that linear control techniques can be applied. This differs entirely from conventional linearization in that feedback linearization is achieved by exact state transformation and feedback, rather than by linear approximations of the dynamics [12]. The idea of feedback linearization, i.e., of canceling the nonlinearities and imposing a desired linear dynamics, can be simply applied to a class of nonlinear systems described by the so-called companion form, or controllability canonical form. A system is said to be in companion form if its dynamics is represented by

$$x^n = f(x) + b(x)u$$

where u is the scalar control input, x is the scalar output of interest, $X = [x, \dot{x}, ..., x^{(n-1)}]^T$ is the state vector, and $f(x)$,b(x) are nonlinear functions of the states. This form is unique in the fact that, although derivatives of appear in this equation, no derivative of the input u is present. Note that, in the state space representation, the above equation can be written

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ \cdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} x_2 \\ \cdots \\ x_n \\ f(x) + b(x)u \end{bmatrix} \tag{1}$$

From systems which can be expressed in the controllability canonical form, using the control input (assuming b to be non-zero)

$$u = \frac{1}{b}[v - f]$$

We can cancel the nonlinearities and obtain the simple input-output relation (multiple integrator form) $x^{(n)} = v$. Thus, the control law $v = -k_0 x - k_1 \dot{x} - ... - k_{n-1} x^{(n-1)}$ with the chosen so that the polynomial $p^n + k_{n-1} p^{n-1} + ... + k_0$ has all its roots strictly in the left half complex plane, leads to the exponentially stable dynamics $x^n + k_{n-1} x^{n-1} + ... + k_0 x = 0$ which implies that $x(t) \to 0$. For tasks involving the tracking of a desired output $x_d(t)$, the control law $v = x_d^{(n)} - k_0 e - k_1 \dot{e} - ... - k_{n-1} e^{(n-1)}$ (where $e(t) = x(t) - x_d(t)$ is the tracking error) leads to exponentially convergent tracking. For the case of helicopter, we should find a state transformation to rewrite the equations in the form of (1). We use the input to state linearization method [12].

There is a input to state linearization method for SISO systems in [12]. We apply this method with some changes to the helicopter model which is a MIMO system. Suppose that the vector $z(x) = [z_1(x)\ ...\ z_n(x)]^T$ is the new state vector, called linearized state vector. To obtain this vector, we should make the following equations

$$\nabla z_1 a d_f^i g = 0 \quad i = 0, ..., n-2 \quad \textit{and} \quad \nabla z_1 a d_f^{n-1} g \neq 0 \tag{2}$$

where is the Lie Bracket of f and g , and is a third vector field defined by $ad_f g = \nabla g \quad f - \nabla f \quad g$.

According to these equations, we can find $z_1(x)$. The linearized vector $z(t)$ is

$$z(x) = \begin{bmatrix} z_1 & L_f z_1 & \dots & L_f^{n-1} z_1 \end{bmatrix}^T \tag{3}$$

where $L_f z_1$ is the Lie Derivative of f and z_1 , and defined by $L_f z_1 = \nabla z_1 \, f$.

The outputs of the system are x_1, x_4 . We separate the states of the system into two sections ($\begin{bmatrix} x_1 & x_2 & x_3 & | x_4 & x_5 \end{bmatrix}^T$), and use the first part to control the first output by the first control input, and the same for the second part.

We start to find $z_1(x)$ for the first set of states of the helicopter model. In this part of the model $n = 3$, so we have the following equations

$$\frac{\partial z_1}{\partial z_i} = 0 \quad for \quad i = 2,3 \qquad , \qquad \frac{\partial z_1}{\partial z_i} \neq 0$$

From these equations, we define $z_1 = x_1$. We use this method to linearize half of the system. It means that we linearized first three equations of the model by the first control input, and the last two equations by the second one. Thus, we obtain the states according to the (3)

$$\begin{aligned} z_2 &= L_f z_1 = X_2 \\ Z_3 &= L_f^2 z_1 = a_0 + a_1 x_2 + a_2 x_2^2 + (a_3 + a_4 x_4 - \sqrt{a_5 + a_6 x_4}) x_3^2 \end{aligned}$$

According to [12], we have

$$u_1 = \alpha(x) + \beta(x) v, \qquad \alpha(x) = -\frac{L_f^n z_1}{L_g L_f^{n-1} z_1}, \beta(x) = \frac{1}{L_g L_f^{n-1} z_1}$$

For the first three states, we have $n = 3$. The first control input has been chosen so that the characteristic equation of the first output becomes $(s+30)(S+3)^2 = 0$. For the last two states, we define the second control input to cancel the nonlinearity of the fifth equation. The characteristic equation of the second output is $(s+1)^2 = 0$.

4 BELBIC Controller

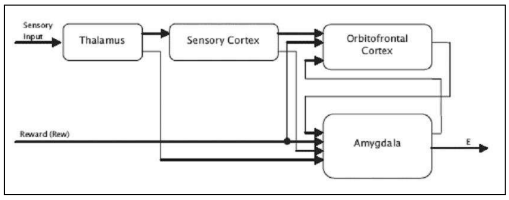


Figure 1: Computational model of emotional learning in the Amygdala

BELBIC is an abbreviation for Brain Emotional Learning Based Intelligent Controller. Motivated by the success in functional modeling of emotions in control engineering applications, a structural model based on the limbic system of mammalian brain, for decision making and control engineering applications has been developed. The computational model of emotional learning in the amygdala, based on Moren and Balkenius model, is depicted in Figure 2 [13, 14]. The main parts that are responsible for performing the learning algorithms are orbitofrontal cortex and amygdala.

BELBIC controller has some sensory inputs. One of the designer's tasks is to determine the sensory inputs. This controller has two states for each sensory input. One of these two is amygdala's output and another is the output of orbitofrontal cortex. Therefore, the number of sensory inputs has a key role in BELBIC controller. Usually the sensory inputs are rich signals [10].

Consider the i-th sensory input as s_i . Then we have amygdala and orbitofrontal cortex outputs

$$A_i = s_i v_i$$

$$O_i = s_i w_i$$

v, w are two states for the related sensory input. These states will be updated by the following equations

$$\triangle v_i = \alpha.s_i.\max(0, rew - \sum A_i),\ \triangle w_i = \beta.s_i.(rew - \sum A_i - \sum O_i - \max(S_i))$$

where α and β are training coefficients. We have a function named Reward. This function has a great role in BELBIC controllers. Reward is like its name. The controller strives to increase this reward. Therefore, the designer must define a reward function that has its maximum values in the most desired regions. This reward function could be frequency domain function or a normal mathematic function.

Amygdala acts as an actuator and orbitofrontal cortex acts as a preventer. Therefore the control effort of BELBIC controller is

$$u = \sum A_i - \sum O_i$$

BELBIC is a controller that has only one output. Therefore for systems with more than one control inputs we must use one BELBIC controller for each control input. As it can be seen there are several tuning parameters for each sensory input. The general algorithm for tuning these parameters is trial and error.

We use continuous form of BELBIC in this paper. In continuous form, the BELBIC states are updated with not a discrete relation but a continuous one. These continuous relations are

$$\dot{v}_i = \alpha.s_i.(rew - A_i),\ \dot{w}_i = \beta.s_i(rew + s_i + O_i - A_i)$$

To control the helicopter in take-off and landing problem, we have two references to be tracked. One of them is desired height and the other one is the desired value of collective pitch angle of rotor blades. We must design a BELBIC controller for each control input of helicopter. For throttle we have one sensory input and it is height's error.

$$s_1 = h_d - h$$

The reward function for this BELBIC controller is as figure 2.

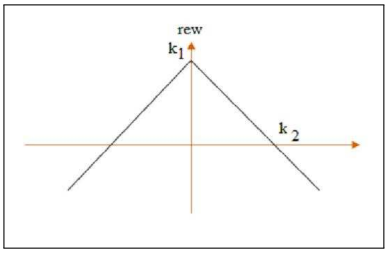


Figure 2: Reward Function

The reward function's parameters are positive real numbers. As it could be seen, the BELBIC controller receives the maximum reward when the sensory input is zero. According to the fact that the sensory input is an error signal, the BELBIC controller tries to vanish the sensory input and it means tracking. With this reward function we have reward and punishment together for BELBIC controller. There are some areas that reward takes negative values. It could help the controllers to sense whole domain of sensory inputs.

The training coefficients and reward function's parameters for the first BELBIC controller are as follows:

$$\alpha = 5, \beta = 34, k_1 = 1000, k_2 = 100$$

The BELBIC controller for servo mechanical control has a sensory input and it is the error of collective pitch angle of rotor blades.

$$s_i = \theta_d - \theta$$

The BELBIC controller for the collective servomechanism control input is like the one for throttle control and with the same reward function but with different parameters. The training coefficients for the first BELBIC controller are as follows

$$\alpha = 3,\ \beta = 1.8,\ k_1 = 1200,\ k_2 = 80$$

5 Simulation Results

To see the performance of the two mentioned controllers, we have simulated the controlled system in Simulink. The height of the helicopter should track a desired path. The second output has a set point too. A sinusoidal path has been considered for the height, and second output kept constant. It can be seen from these simulations that the tracking performance of BELBIC controller for the height is better than Feedback linearization controller. The simulation results of controller system by BELBIC controller and feedback linearization controller have been demonstrated in the figures 3, and figure 4 demonstrates the control inputs of the system.

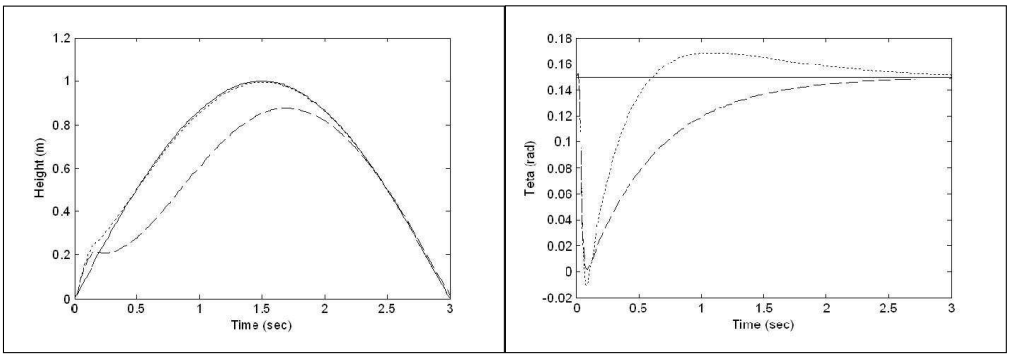


Figure 3: : Height (left) and Collective rotor blade angle (right) of helicopter (*Solid*: set point, *Dashed*: feedback linearization, *Dotted*: BELBIC)

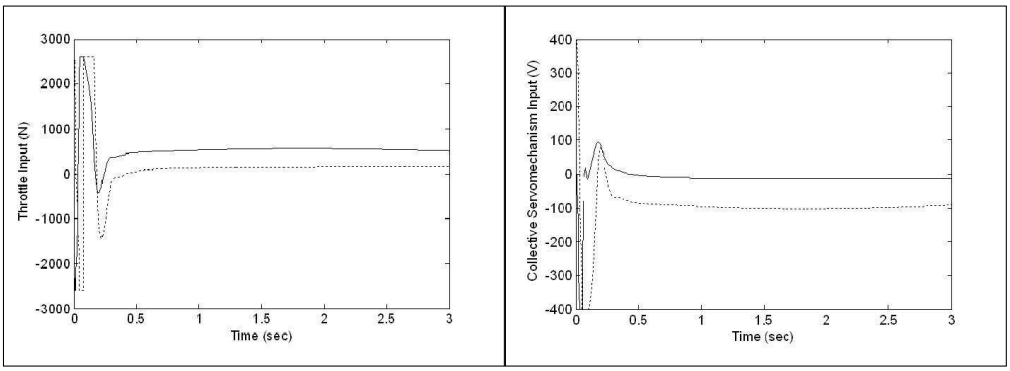


Figure 4: First (left) and second (right) control input of helicopter (*Solid*: feedback linearization, *Dotted*: BELBIC)

6 Summary and Conclusions

In this paper a BELBIC controller and a feedback linearization technique applied to a nonlinear model of a helicopter to attain path tracking of a given path for the height of the helicopter. The system has five states and two control inputs. A continuous time BELBIC controller has been applied to the system, and tuned for the best performance. An input to state linearization has been used to control the system too. In this method the states of the system have been separated into two parts, and each part has been controlled by one of the control inputs. The transient response of the BELBIC controller is better than the case with feedback linearization controller, but in the sense of steady state the performance of both controllers is good. However, there is an important disadvantage for this controller that is stability guarantee. The BELBIC is a very suitable method to solve control engineering problems, and it can be applied to a large variety of the linear and nonlinear systems.

References

[1] W. Garrad, And E. Low, "Eigenspace design of helicopter flight control systems", Technical report, Dept. of Aerospace Engineering and Mechanics, University of Minnesota, Nov. 1990.

[2] K. Kienitz, Q. Wu, And M. Mansour, "Robust stabilization of a helicopter model". Proceedings of the 9th CDC, pp. 2607-2612, 1990.

[3] K. Furuta, Y. Ohyama, And Yamanao, "Dynamic of RC helicopter and control", in "Mathematics and computers simulation XXVI"(North Holland, Amsterdam, 1984), pp. 148-159.

[4] Tj. Palle'It, B.J. Wolfert And S. Ahmad, "Real time helicopter flight control test bed". Technical Report TR-EE 91-28, School of Electrical Engineering, Purdue University, 1991.

[5] H. Sira-Ramirez, M. Zribi, S. Ahmad, "Dynamical sliding mode control approach for vertical flight regulation in helicopters", IEE Proc.-Control Theory Appl., Vol. 141, No. I, January 1994.

[6] J. Kaloust, C. Ham, Z. Qu, "Nonlinear autopilot control design for a 2-DOF helicopter model", IEE Proc.-Control Theory Appl., Vol. 144, No. 6, November 1997.

[7] Sharbafi A. Maziar, Lucas Caro, Mohammadinejad Aida, Yaghobi Mostafa, "Designing a Football Team of Robots from Beginning to End", International journal of Information Technology, Vol 3, No. 2, 2006.

[8] C. Lucas, S. Moghimi, "Applying BELBIC (Brain Emotion Learning Based Intelligent Controller) to an Auto landing System", Conference: WSEAS AIKED'03, 2003.

[9] Tutunchi Ali Ghasem, "Optimal BELBIC control with application to Autopilot control", Master's Thesis, University of Tehran, 2006.

[10] C. Lucas, D. Shahmirzadi, N. Sheikholeslami, "Introducing BELBIC: Brain Emotional Learning Based Intelligent Controller", International Journal of Intelligent Automation and Soft Computing, Vol. 10, No. 1, pp. 11-22, 2004.

[11] G. Meyer, R.L. Hunt And R. Su, "Design of a helicopter autopilot by means of linearizing transformations", Guidance and Control Panel, 35th Symposium, AGARD CP321, Paper 4, 1983.

[12] Jean-Jacques E. 12, Weiping Li, *Applied Nonlinear Control*, Prentice Hall, 1991.

[13] J. Moren, C. Balkenius, "A Computational Model of Emotional Conditioning in the Brain", in Proc. workshop on Grounding Emotions in Adaptive Systems, Zurich, 1998.

[14] J. Moren, C. Balkenius, "A Computational Model of Emotional Learning in The Amygdala: From animals to animals", in Proc. 6th International conference on the simulation of adaptive behavior, Cambridge, Mass., The MIT Press, 2000.

Saeed Jafarzadeh, Rooholah Mirheidari, Mojtaba Barkhordari
Iran University of Science and Technology
Department of Electrical Engineering
Address: Daneshgah Street, Hengam Street, Resalat Sq., Tehran, Iran
E-mail: sjafarzadeh, rmirheidari@ee.iust.ac.ir, mbarkhordary@iust.ac.ir

Mohammad Reza Jahed Motlagh
Iran University of Science and Technology
Department of Computer Engineering
Address: Daneshgah Street, Hengam Street, Resalat Sq., Tehran, Iran
E-mail: jahedmr@iust.ac.ir

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 343-348

Designing PID and BELBIC Controllers in Path Tracking Problem

Saeed Jafarzadeh, Rooholah Mirheidari, Mohammad Reza Jahed Motlagh, Mojtaba Barkhordari

Abstract: This paper proposes a new intelligent control approach for path tracking of a vehicle used in automated highway systems. Brain emotional learning based intelligent controller (BELBIC) is an intelligent controller based on the model of emotional part of brain. A modified BELBIC controller has been applied to a sixth order model of the vehicle which should track any normal path. A model with the coupling terms between steering angle and traction force is considered. The simulation results of this controller has compared with a PID based controller. The policies for PID based and BELBIC controller are the same. Controller is designed for vehicle to accelerate according to distance from reference point and to steer according to direction of vector which is defined from vehicle to reference point. According to simulation results it can be seen that the performance of tracking improved using BELBIC controller.
Keywords: Path Tracking, BELBIC, Automated Highway Systems.

1 Introduction

Normally, vehicles are manually controlled. But there are situations where manual control could be replaced by autonomous control. Sometimes it has to be done because of polluted environments such as chemical factories and nuclear power stations. Also in AHS (Automated Highway Systems) and ITS (Intelligent Transport Systems) researches, the car is needed to track a designated path which is defined by central guide controller [1, 2].

A lot of work has been done on the path tracking control of vehicles [3, 4, 5]. Some useful controllers have been proposed [6, 7]. In these papers writers have been ignored the coupling terms using small steering angle assumption. This assumption may help to simplify the controller, but may lead to unstable situations. Therefore a model with the coupling terms between steering angle and traction force is considered.

In this paper the BELBIC method applied to a relatively complicated model of car to force it to track a given path. The BELBIC controller is an intelligent controller which could be used in many fields [8, 9]. The powerful aspect of this controller is laid in a function named reward. This reward function is the basic logic of this controller. In path tracking problem the car should follow a moving point. The task is also has been done using PID controllers. Then two methods compared and the benefits of using each of these controllers shown.

At first the considered model is shown. After that the PID controller is used for path tracking problem and in next section BELBIC controller introduced briefly. The simulation results will be illustrated then and after all the comparisons made to obtain a good conclusion.

2 Car Dynamics Model

A model consists of coupling between traction force and steering angle used in this paper. The following model has this specification [10]:

$$\dot{u} = vr - fg + \frac{u^2(fk_1 - k_2)}{M} + c_f\frac{v+ar}{Mu}\delta + \frac{r}{M} \quad (1)$$

$$\dot{v} = ur - \frac{(c_f + c_r)v}{Mu} + \frac{(bc_r - ac_f)r}{Mu} + \frac{c_r + T}{M}\delta \quad (2)$$

$$\dot{r} = -\frac{(ac_f - bc_r)v}{l_z u} - \frac{(b^2c_r + a^2c_f)r}{l_z u} + \frac{a(c_f + T)}{l_z/\delta} \quad (3)$$

$$\dot{x} = u\cos\varphi - v\sin\varphi \quad (4)$$

$$\dot{y} = u\sin\varphi + v\cos\varphi \quad (5)$$

$$\dot{\varphi} = r \quad (6)$$

This model describes vehicle dynamics as a sixth order state space model with two inputs which are traction force and steering angle. It is considered that all the states of vehicle could be measured.

The control objective in this paper is to guide vehicle to track the given path. Therefore the system's outputs are lateral and longitudinal position of vehicle. In general view of the system, it is a system with two inputs, two outputs, and two references.

In this model **u** is longitudinal velocity, **v** is lateral velocity, **x** is longitudinal position, **y** is lateral position, and φ, r are vehicle's angle and its rate respectively. The inputs of vehicle are T, δ which are traction force and steering angle of vehicle respectively. Other parameters are constants that have been illustrated in Table 1.

Parameter	Value	Parameter	Value	Parameter	Value
M	1480 kg	f	0.02	c_r	95 kN/rad
h	0.35 m	a	1.05 m	k_1	0.005 $N.s^2/m^2$
l_z	2350 $kg \cdot m^2$	b	1.63 m	k_2	0.41 $N.s^2/m^2$
g	9.81 m/s^2	c_f	135 kN/rad		

Table 1: Parameters of the vehicle model

3 PID Control Method

To control the mentioned system, PID controllers could be used. Policies will be defined for controllers which could help us achieving the control objective. These policies are derived from physical intuition of system. Consider a vector connecting the vehicle to desired point. The controller will try to maintain the vehicle's direction in the vector's direction. To do so controller has to minimize the difference signal between the vehicle's angle and the vector's angle. This policy shows us the way to find a good steering angle. In fact this difference signal should be regulated. PID controllers could be used to do that. But this signal has to be built with state feedbacks.

The other policy is to maintain the traction force in relation with the defined vector length. In fact this vector's length is the distance between vehicle and desired point. It's obvious that this also means a tracking problem which could be solved using PID controllers. Therefore the angle and length of the defined vector's signals has to be built using feedback and reference signals. Then the error of angle and speed could be made and use in controllers.

Let's consider $x_c \cdot y_c \cdot x_d \cdot y_d$ the lateral and longitudinal positions of vehicle and desired point respectively. Then the desired speed defined as:

$$s_d = 1 + 0.1\sqrt{(x_d - x_c)^2 + (y_d - y_c)^2} \tag{7}$$

It could be seen that this desired speed is in relation with the length of connecting vector from vehicle to desired point. It means that vehicle's speed increases when the distance increases and it's logically true. Therefore the input for PID controller of traction force is $(s_d - u)$.

Let's define the α as the angle of defined vector. This angle could be obtained using lateral and longitudinal errors of position. The input for PID controller of steering angle is $(\alpha - \varphi)$.

The Ziegler-Nicholes technique used to find the PID controllers' parameters. This controller is based on classic control and the vehicle could track the path using this controller.

4 BELBIC Controller

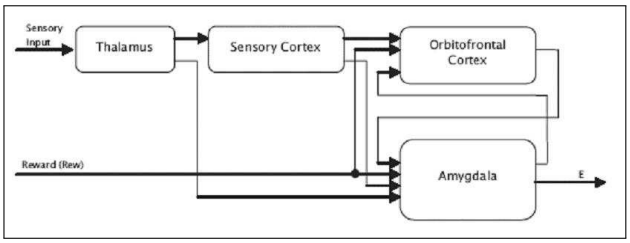


Figure 1: Computational model of emotional learning in the Amygdala

BELBIC is the abbreviation for brain emotional learning based intelligent controller. Motivated by the success in functional modeling of emotions in control engineering applications, a structural model based on the limbic system of mammalian brain, for decision making and control engineering applications has been developed [11]. The computational model of emotional learning in the amygdala, based on Moren and Balkenius model [12], is depicted in Figure 2. The main parts that are responsible for performing the learning algorithms are orbitofrontal cortex and amygdala.

BELBIC controller has some sensory inputs. One of the designer's tasks is to determine the sensory inputs. BELBIC controller has two states for each sensory input. One of these two is amygdala's output and another is the output of orbitofrontal cortex. Therefore the number of sensory inputs has a key role in BELBIC controller. Usually the sensory inputs are rich signals [13].

Consider the i-th sensory input as s_i . Amygdala and orbitofrontal cortex outputs are as follows:

$$A_i = s_i v_i \tag{8}$$
$$O_i = s_i w_i \tag{9}$$

v, w are two states for the related sensory input. These two will be updated by following equations:

$$\triangle v_i = a \cdot s_i \cdot \max(0, rew - \sum A_i) \tag{10}$$
$$\triangle w_i = \beta \cdot s_i \cdot (rew - \sum A_i - \sum O_i - \max(s_i)) \tag{11}$$

In these equations α and β are training coefficients. In BELBIC controller there is a function named Reward. This function has a great role in BELBIC controllers. Reward is like its name. The controller strives to increase this reward. Therefore the designer must define a reward function that has its maximum values in the most desired regions. This reward function could be either a frequency domain function or a normal mathematical function.

Amygdala acts as an actuator and orbitofrontal cortex acts as a preventer. Therefore the control effort of BELBIC controller is:

$$u = \sum A_i - \sum O_i \tag{12}$$

BELBIC is a controller that has only one output. Therefore for systems with more than one control inputs designer must use one BELBIC controller for each control input. As it can be seen there are several tuning parameters for each sensory input. The general algorithm for tuning these parameters is trial and error.

The continuous form of BELBIC has been used in this paper. In continuous form the BELBIC states are updated with not a discrete relation but a continuous one. These continuous relations are:

$$\dot{v}_l = a \cdot s_i \cdot (rew - A_i) \tag{13}$$
$$\dot{w}_l = \beta \cdot s_i \cdot (rew + s_i + O_i - A_i) \tag{14}$$

To control the vehicle in path tracking problem the similar policies as in PID based control has been taken. It means that there are two signals to be tracked. One of them is reference point's speed and the other one is defined vector angle.

A BELBIC controller has to be designed for each control input of vehicle. For traction force one sensory input considered and it's speed's error.

$$s_1 = s_d - u \tag{15}$$

The reward function for this BELBIC controller is as figure 2.

The reward function's parameters are positive real numbers. As it could be seen the BELBIC controller receives the maximum reward when the sensory input is zero. According to the fact that the sensory input is an error signal, the reward function drive the BELBIC controller to vanish the sensory input and it means tracking. This reward function has reward and punishment together. There are some areas that reward takes negative values. It could help the controllers to sense whole domain of sensory inputs.

BELBIC controller for steering angle has two sensory inputs. One is angle error and one is defined vector's angle.

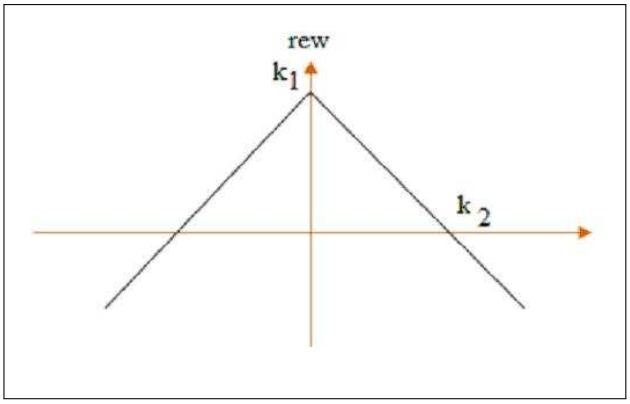


Figure 2: Reward Function

$$s_1 = \alpha - \varphi \quad (16)$$

$$s_2 = \alpha \quad (17)$$

The first sensory input is the one which BELBIC acts on. But BELBIC use the second one to choose its output sign. In fact this BELBIC controller is a BELBIC controller like the one for traction force and with the same reward function with first sensory input that multiply its output with the sign function of the second sensory input. The reward function has been chosen like first BELBIC because their sensory inputs are error signals and therefore the peak of reward function should be happened when sensory input is zero.

5 Simulation Results

To see and compare the performance of the two mentioned controllers the vehicle simulated in a normal path. The results of tracking are shown in figure 3.

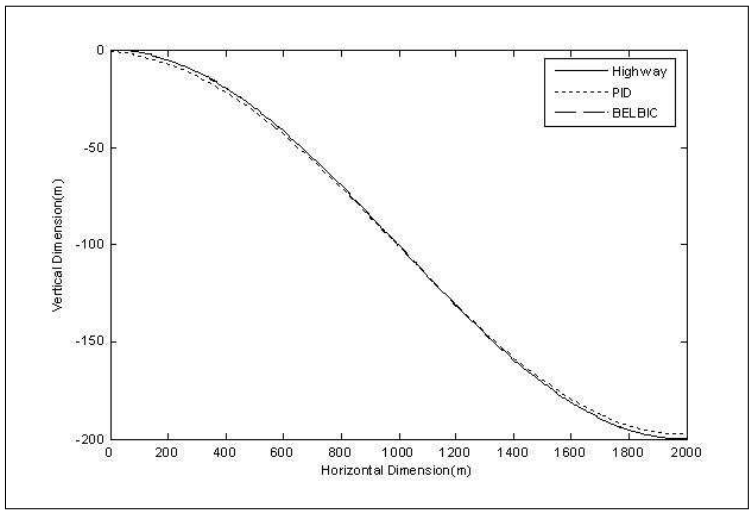


Figure 3: Highway and vehicle controlled by PID and BELBIC

As shown in figure 3 the vehicle tracked the highway perfectly when is controlled by BELBIC. The BELBIC controlled vehicle's trajectory and highway are not distinguishable in figure 3. It can be seen from these simulations that the tracking performance of BELBIC controller is better than PID controller. It's not that surprising because the fact that BELBIC is an intelligent controller.

The traction force and steering angle of both controllers are shown in figure 4 and 5.

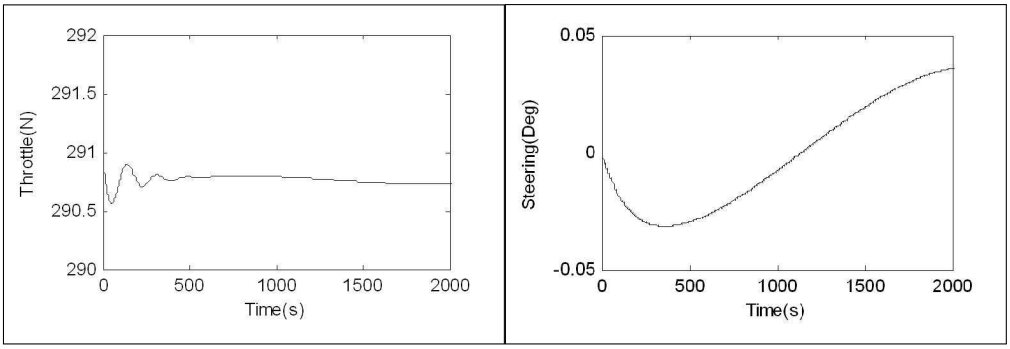


Figure 4: Traction force (left) and steering angle (right) generated by PID controller

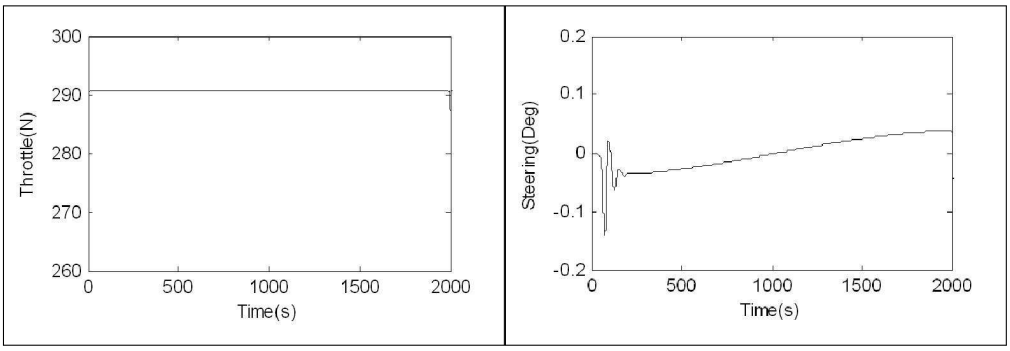


Figure 5: Traction force (left) and steering angle (right) generated by BELBIC controller

As it could be seen from figure 4 and 5 the control efforts are in reasonable range and don't exceed the physical limits.

6 Conclusions

In this paper a BELBIC controller and a PID based controller applied to a complicated model of a vehicle to control the considered vehicle through path tracking problem. There are some modifications made in BELBIC controller to reach a continuous form. The main idea of controllers' design mentioned as two policies and finally the simulation results illustrated to show the advantages and disadvantages of these two controllers.

The maximum deviation of vehicle controlled by BELBIC from desired position is efficiently lower than the PID controlled one, especially in highways with high turn rate. But the control efforts generated by PID controller are less fluctuating in same case, though there is no violation in control efforts and their rates neither in BELBIC nor in PID controller.

The main advantage of both controllers is that they can track any path with normal turn rate.

The number of states which their feedback is needed is same in both controllers, and it's because that the control policy in both controllers is same.

7 Future Works

To enhance theoretic aspect of this work the investigation of the stability conditions of controller could be done. There is no specified way for proving stability of BELBIC controllers; therefore the lyapunov theory could be useful in this field.

References

[1] S. Tsugawa, "Control Algorithms for Automated Driving Systems", Journal of JSAE, Vol. 52, No. 2, pp. 28-33, 1998.

[2] T. Fujioka, M. Omae, "Overview of Control for Automatic Driving System", Journal of the Robotics Society of Japan, Vol. 17, No. 3, pp. 18-23, 1999.

[3] P. Varaiya, "Smart Cars on Smart Roads: Problems of Control," IEEE Transactions on Automatic Control, Vol. 38, No.2, pp. 195-207, 1993.

[4] J. Ackeimann, et ai, "Linear and Nonlinear Controller Design for Robust Automatic Steering," IEEE Transactions on Control Systems Technology, Vo1.3, No.1, pp. 132-143, Mar. 1995.

[5] W. L. Nelson and 1. J. Cox, "Local path control for an autonomous vehicle," in Pruc. 1988 IEEE Int. Cotif: Rohor. and Autoniat., pp. 1504-15, 1988.

[6] Hatipoglu Cem, Ozgiiner Omit, and Unyelioglu A. Konur, "Advanced Automatic Lateral Control Schemes for Vehicles on Highways", Proceedings of the 13th IFAC, San Francisco, USA, pp. 477-482, 1996.

[7] O. J. Sordalen and C. Canudas de Wit. "Exponential control law for a mobile robot: Extension to path following", IEEE Transaction on Robotics and Automation, Vol. 9, Issue 6, PP. 837-842, Dec. 1993.

[8] A. Sharbafi, C. Lucas , "Designing a Football Team of Robots from Beginning to End " , INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY, VOLUME 3, NUMBER 2, ISSN 1305-2403, 2006.

[9] H. Rouhani, R. M. Milasi and C. Lucas, "Speed Control of Switched Reluctance Motor (SRM) Using Emotional Learning Based Intelligent Adaptive Controller," 5th IEEE International Conference on Control and Automation, ICCA'05, Budapest Hungary, June 26-29, 2005.

[10] A. Alloum, "Modelisation et commande dynamique d'un vthicule pour la stcuritt de conduite," Ph. D thesis, U. T. C, Compiegne, France, 1994.

[11] J. Moren, C. Balkenius, "A Computational Model of Emotional Conditioning in the Brain". in Proc. workshop on Grounding Emotions in Adaptive Systems, Zurich, 1998.

[12] J. Moren, C. Balkenius, "A Computational Model of Emotional Learning in The Amygdala: From animals to animals". in Proc. 6th International conference on the simulation of adaptive behavior, Cambridge, Mass., The MIT Press, 2000.

[13] C. Lucas, D. Shahmirzadi, N. Sheikholeslami, "Introducing BELBIC: Brain Emotional Learning Based Intelligent Controller". International Journal of Intelligent Automation and Soft Computing, Vol. 10, No. 1, pp. 11-22, 2004.

Saeed Jafarzadeh, Rooholah Mirheidari, Mojtaba Barkhordari
Iran University of Science and Technology
Department of Electrical Engineering
Address: Daneshgah Street, Hengam Street, Resalat Sq., Tehran, Iran
E-mail: {sjafarzadeh, rmirheidari}@ee.iust.ac.ir, mbarkhordary@iust.ac.ir

Mohammad Reza Jahed Motlagh
Iran University of Science and Technology
Department of Computer Engineering
Address: Daneshgah Street, Hengam Street, Resalat Sq., Tehran, Iran
E-mail: jahedmr@iust.ac.ir

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 349-352

E-Learning Using the Basic Knowledge Management Process in the Organizational Growth

Viorina-Maria Judeu, Emma-Margareta Văleanu

Abstract: The educational needs of modern virtual communities like company teams and student in an educational institution are explored and possible solutions integrating e-learning capabilities and advantages of knowledge management process are presented. Basic knowledge management processes are reviewed and most important functionalities of is discussed. Influence of knowledge management processes in development of new forms of advance learning is described and advantages of electronic learning (e-learning) application in the knowledge management process in an organization or institution in order to facilitate organizational success and growth.
Keywords: collaboration, knowledge management, e-learning, knowledge management process.

1 Introduction

The educational needs of modern virtual communities like company teams and student in an educational institution are explored and possible solutions integrating e-learning capabilities and advantages of knowledge management process are presented. Basic knowledge management processes are reviewed and most important functionalities of is discussed. Influence of knowledge management processes in development of new forms of advance learning is described and advantages of electronic learning (e-learning) application in the knowledge management process in an organization or institution in order to facilitate organizational success and growth. At the end of research common features of both filed are defined. Future research efforts will be dedicated on better and more effective integration of knowledge management capabilities in e-learning delivery and powerful use of learning materials and activities in the process of knowledge manipulation and exchange in organizations and institution in order to provide organizational success and prosperity.

2 Learning Strategies

Among the learning strategies that underlie learning object structural models, generative learning has special importance given the specificity of the learning environment in which it may be applied: an environment comprised of highly-motivated students whose knowledge background is basically homogeneous and of a high level, and who are aware of the need to work together to find the solution to a problem and with the purpose of qualitatively improving the activity of the entire group, not just each individual student. This is the situation characteristic of those wide-spread and common situations in which, due to structural growth or re-engineering of procedures, there is a need for training within the workplace organization, training occurring between co-workers and not led by an instructor. Users involved in a learning process in which they must cooperate in identifying the problems to be tackled and resolved, the themes to be handled in a self-updating process, the modes of learning interaction and action and, as a result, shared "meta-cognition", will configure their activity, in strategic terms, as generative learning.

This strategy, when it expands to include learning tools based on the use of digital and online technologies (as a rule: learning is always among the top sectors that most rapidly assimilates innovations in communications technology), becomes so closely identified with knowledge management that the two processes cannot be distinguished from each other. In fact, if we take into consideration corporate knowledge management as an activity aimed at enhancing the knowledge and skills within the work environment through the presentation and sharing of tacit knowledge and the organization of shared or explicit knowledge (informational and documental patrimony), the overlap between this activity and generative learning is clear, both from a methodological standpoint and, consequently, from a technological one in the choice and implementation of suitable tools of interaction.

3 Processes of Knowledge Management and the Learning Process

The most common used process of knowledge manipulation is capturing, storage and distribution of knowledge. This sequence is part of the training in each successful organization or institution. People use different types of repositories and specialists implement different technologies for organization of knowledge collecting, storage and delivery on demand. Purpose of the process is to improve qualification of team members and this way achievement of better results. Basic knowledge processes defined by Nonaka and Takeuchi ([1, Nonaka,Takeuchi]) are socialization, externalization, internalization and combination and their implementation in the transfer of tacit knowledge.

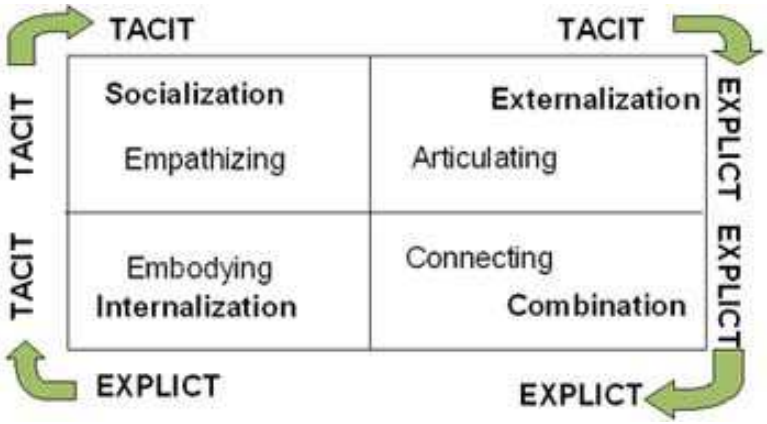


Figure 1: The basic processes defined by Nonaka and Takeuchi ([9])

Frappaolo and Toms define another classification of knowledge management processes ([2, Frappaolo, Toms]):

1. Socialization: Transfer tacit knowledge from one person to another person
2. Externalization: Translate tacit knowledge into explicit knowledge in a repository
3. Combination: Combine different bodies of explicit knowledge to create new explicit knowledge
4. Internalization: Extract the explicit knowledge from a repository that is relevant to a particular person's need and deliver it to that person where it is translated into tacit knowledge
5. Cognition: Apply tacit knowledge to a business problem

Simple development and delivery of learning resources can not satisfy requirements of information society of professionals skills achievement, knowledge sharing and exchange and gaining competencies in specific domains of science and real life necessary for individuals and organizational success and prosperity. That is way education has to be a process of sharing and acquirement of knowledge, skills and competencies. Advantage of Knowledge management is very useful for that process. Knowledge management is indivisible part of teams training so capturing of knowledge process is very similar to the processes related to selection of most appropriate learning content in e-learning.

At high level of understanding the desired outcome of learning should be knowledge acquisition and in combination with some practical skills gained in the process of education they have to present some type of competence. This competence learner should be able to apply in his professional duties and to execute the task correctly. So education seems to be very important part of development of successful team. Learning on demand or just -in-time training is very appropriate forms of education at work.

Activities involved in standard education have to be implemented in different types of trainings delivered at organizations and institutions. This way communication and collaborative will be improved and free exchange of competencies will be provided.

External lectures can heighten level of expertise in a team and practical exercises involved in trainings can give workers more experience and skills in execution of their professional duties.

Project development tools and capabilities of corporate (organizational) system are critical for execution of different team and individual tasks and delivery of necessary information, data and document in time. Use of

project management tools is very useful in the process of education. It allows projects developed by students to be scheduled and implemented on time. Another advantage of their integration in education is students get used to work with the tools and they know which are strong point and drawback of project management tolls implementation in their work and how to manage existing problems related to their use.

Collaboration is very critical process for each of both activities. Means for communication and collaborations are one of the most important characteristics of successful education and team work. They could include synchronous and asynchronous communication and different tools related to work in groups or different types of virtual communities. In education students and teachers have to exchange information related to learning activities or specific topics of the proposed learning content. Participant in education involved in different types of groups have to exchange knowledge, skills and competences. Team members in an organization or institution have to send and receive important information or data related to their duties. Free exchange of knowledge and data and capabilities for collaborative editing of documents become even more critical when different members of the team are at distance (different offices, cities or countries).

Possibility of help desk support and delivery of information on demand is very powerful feature of each supporting system. It allows different types of problems to be decided professionally and as soon as it is possible. That eliminates pressure and confusion of users and workers and decrease time necessary for a problem solution n and task execution.

Paper focuses on basic characteristics of e-learning and knowledge management and main task of the research is to find common features of both domains. Combination of advantages of both domains facilitates delivery of high quality education for satisfying specific educational needs of team members ([3, Novak, Gowin]).

Management of the cognitive context-understood as the reconstruction of the fabric of concepts and relations, representation in the form of a concept map and management of interactive functions that are inherent to or can be situated within the context itself-is a significant exigency both within the learning environment, and in particular for e-learning, as well as within the corporate environment as part of knowledge management processes.

In order to take advantage of knowledge mapping and the functions connected with it, software tools capable of supporting and managing interaction in a network environment must be designed and created on the basis of guidelines aimed at pursuing not only the ergonomics, efficiency and efficacy of use, but also flexibility and completeness during the phases of construction, updating and analysis of the relational fabric.

Starting from these premises, it is possible to design the technological features of a tool for online interaction configured to support generative learning and knowledge management and therefore with a two-fold possibility for fruition, and aimed both at sharing and contextualizing tacit and explicit knowledge to both guide and to analyze the context under examination in order to provide a decisional support tool ([4, Novak]).

There are four axes along which re-engineered learning activity runs, from a purely strategic standpoint, through a design approach that takes into consideration technological innovation (learning activity in general, and not e-learning or blended learning organized as online training plus short attended seminars, because there is no question of the potential for applying the results of this re-engineering including in environments in which learning is primarily traditional in nature, i.e., normal, on-going classroom lessons). Added to the first two-content transmission and interaction between stakeholders and the training process, both of which are clear and therefore do not require further comment-are control of the training process (an axis which in online training takes the form of tracking activities, reports of the data collected and portfolio elaboration) and context management.

Here, we will concentrate on this last axis. By "context management" is meant the set of functions that permits the reconstruction of the knowledge context (a subject-based knowledge in formal and non-formal learning environments that becomes a patrimony of know-how, knowledge and skills within organizations and informal training environments) through graphic tools and on an individual and collaborative basis, its representation and use from a documental basis and as hermeneutic support, its availability on an ontological basis as a virtual site for negotiating, its use-value as a tool for navigation and progressive exploration of knowledge and, finally, its support role for the structured placement of documentary resources and asynchronous interaction around the topics examined.

From this definition, we can deduce how functionally rich a context management technological tool must be and, as a result, how complex it is to design and create one.

Nonetheless, the basic aspect of the context management support tool, both technologically and from a structural basis, is how to portray the context in the form of a concept map that is advanced both in terms of a dynamic reconstruction of a digital document, as well as offer a sophisticated and codified graphics option to enrich the ability to communicate information that is attributive in nature (size and colors of graphic objects, use of the background as a further element in informational support).

The context management environment, inserted within the operational environment of generative learning or

knowledge management, must, first of all, offer the context structure in the form of a graphic map, or graph, and implement on this map the set of functions listed above, not because all the functions actually reference the organizational form of knowledge given in the map, but also because to assign the map the role of interface between the user and cognitive environment implies the immediate and on-going transmittal of the ontological structure itself which becomes at the same time the entrance threshold for exploration, the schematic and summary document of the content and relational fabric underpinning it and, finally, symbol itself of the context and identifying element ([5, Gonzales,Artiles]).In other words, the interface of the context management environment founded on knowledge mapping takes on its own learning value that goes beyond and is radically different from mere functional support and demands of ergonomics and efficient interaction.

4 Conclusions

This paper has illustrated how e-Learning techniques and technology can be used to enhance knowledge management in an enterprise and provide the benefits of both. From the context management definition we deduced how functionally rich a context management technological tool must be and, as a result, how complex it is to design and create one. Creating models and maps or various scenarios we can provide a roadmap for the evolution of new systems that will provide both the efficient capture of knowledge and the efficient delivery of knowledge.

References

[1] I. Nonaka, H. Takeuchi, *The knowldge-creating company: How Japanese companies create the dynamics of innovation*, New york, Oxford university press, 1995

[2] C. Frappaolo, W. Toms, "Knowledge Management: From Terra Incognito to Terra Firma," *The Delphi Group*, 1997.

[3] J.D. Novak, D.B. Gowin, *Learning how to learn*, New York:Cambridge University Press, 1984.

[4] J.D. Novak, *Learning, creating, and using knowledge: Concept Maps as facilitative tools in schools and corporations*, Mahweh, NJ:Lawrence Erlbaum Associates, 1998.

[5] G. Gonzales, S. Artiles, "Simetria de la tecnica de mapas conceptuales y la dimension informacional de la gestion de conocimiento de la s organisaciones: GECYT como caso de studio," *1st International Conference on Concept Mapping*, Pamplona: Direccion de publicaciones de la Universidad Publica de Navarra, 2002.

[6] H. Garnder, *Intelligence Reframed. Multiple Intelligence for the 21st Century*, Basic Book, Perseus Books Group, 1999.

[7] J.F. Sowa, *Knowledge representation: logical, philosophical and computational foundations*, Brooks Cole Publishing Co., 2000.

[8] R. Davis, H. Shrobe,P. Szolovits, "What is a Knowledge Representation?," *AI Magazine*, 14(1):17-33, 1993.

[9] http://www.allkm.com/km-basics/knowledge-process.php

[10] http://www.jfsowa.com/pubs/semnet.htm, John F. Sowa, Semantic Networks

Viorina-Maria Judeu, Emma-Margareta Văleanu
Agora University
Informatics Economic Department
Oradea, Romania
E-mail: {viorina,evaleanu}@univagora.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 353-357

Issues & Trends in AutoConfiguration of IP Address in MANET

Harish Kumar, R.K. Singla, Siddharth Malhotra

Abstract: Address autoconfiguration provides convenience in implementing Mobile Ad hoc Network (MANET). The topology of the network can change randomly due to unpredictable mobility of nodes. This behavior results in certain issues like partitioning, merging, duplicate address detection, security / authenticity, related to IP address allocation to the mobile nodes. In this paper we have investigated these issues. By collecting the data from IEEE Xplore^{TM} and Springer's "Lecture notes on Computer Science", recent trends have been studied and future research direction has been established.
Keywords: Auto-configuration, Duplicate address detection, MANET, Partition, Merge

1 Introduction

A Mobile Ad-hoc Network (MANET) is a self-configuring network of mobile hosts connected by wireless links. The hosts are free to move randomly and organize themselves arbitrarily. The topology may change rapidly and unpredictably. Due to this nature, it becomes difficult to make use of the existing techniques for network services. The major issues in MANET are routing, multicasting/broadcasting, address autoConfiguration, transport layer management, power management, security, Quality of Service (QoS), products[1].

A technique to autoconfigure IP addresses for MANET nodes is required. Fixed IP address or Dynamic Host Control Protocol (DHCP) cannot be used due to rapid topology change and non-availability of network infrastructure. Proposed techniques can be classified into three categories [2]: *Best effort allocation* schemes cannot guarantee address uniqueness. Prophet[3] proposes a complex address generation function to generate a sequence of addresses to be assigned to new nodes. It may need some mechanism, such as Duplicate Address Detection (DAD) or weakDAD [4], to resolve address conflicts. DAD causes broadcast storm problem and weakDAD introduces overhead. In *Centralized allocation* a server is deployed to manage all addresses. DHCP [5] is an example, but it needs broadcast for server discovery and DAD. A longer periodic broadcast interval can help to reduce overhead, but it also results in longer latencies. In *Decentralized allocation* a host could acquire an address by itself or from a neighbor and then performs DAD to ensure the uniqueness of the address. Host may randomly select an address. In MANETconf [7], each host stores all addresses used in the MANET and a new host acquires an address from one of its neighbors. The neighbor then broadcasts a query, on behalf of the new host, for DAD. In this paper issues and trends related to address auto-configuration are discussed.

2 Issues in Address Allocation

2.1 Partitioning of MANET

The division of a network into two or more sub-networks is known as partitioning. It leads to IP address leak. Partitioning can be of two types: graceful and graceless. If nodes leave after informing their neighbors then it is graceful otherwise graceless. In graceful partition, the newer nodes joining the network can reclaim IP addresses. But graceless partitioning leads to address leakage and there is requirement of some technique to detect it. Figure 1 explains IP address leakage. Before partitioning there is only one MANET with IP address range from 192.168.1.51/24 to 192.168.1.54/24. After partitioning it is divided into two independent MANETs each consisting of some of the IP addresses from the range of original MANET. Addresses in Partition-I can not be assigned to new incoming node in Partition-II until it is not able to detect this partitioning and vice-versa. This leads to decrease in the number of IP addresses that can be allocated.

MANETconf [7] uses a universal identifier for all nodes. Node 'N' which is configured with the lowest IP address represents MANET identifier. When it splits, only one partition own 'N' (let this partition be A and other B). When a new node M arrives in B, 'N' is not able to answer DAD request. Then M's initiator realizes that 'N' doesn't belong to its network. M's initiator then broadcast a message to inform that network has split. Hence, all nodes from B are able to update their address list. The process is similar for A's nodes. Addresses of B partition become free address for A and vice versa. ZAL [6] assumes that partitioning is always graceful.

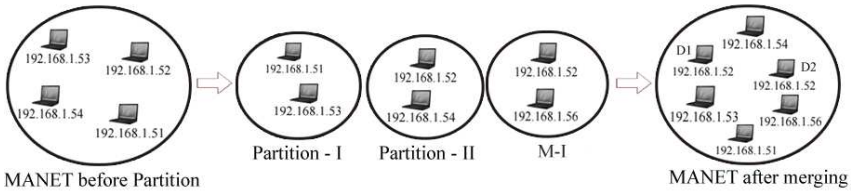


Figure 1: Partitioning & Merging of MANET

2.2 Merging of several MANETs

Combining of two or more networks into one bigger network is called merging. It occurs when independent networks come into range of each other. It can cause IP address conflict. Figure 1 explains merging also. Partition-I, Partition-II & M-I are independent networks using their own range of IP addresses. But if these networks merge then few nodes may have same addresses as nodes D1 & D2. To overcome this problem, DAD is required. In MANETconf[7] two nodes that initiate a communication, exchange their identifier. If identifiers are different, then they realize that their networks have merged. Then they act as configured initiators and start reconfiguration of nodes with conflicting address in their own network. In ZAL[6] nothing needs to be done when networks were part of the same larger network because address spaces at different sub-networks were disjoint. Partition ID is used to find out that they belong to the same larger network. If merging networks never met before, ZAL proposes to convert addresses of nodes in smaller networks to that of larger networks. Only addresses in one of the networks can be preserved. The others have to convert. It is a gradual process in which first nodes at the boundaries of smaller networks and then slowly innermost are converted. It is desirable to minimize overhead by minimizing number of address conversions based on distributed algorithms.

2.3 Duplicate Address Detection (DAD)

DAD is required when either a new node joins a MANET or independent networks merge. A new node picks up a tentative IP address. DAD process determines whether this address is available or not. All the nodes having a valid IP address participate in DAD to protect their IP address being used accidentally by new node. The uniqueness check is based on sending a Duplicate Address Probe (DAP) and expecting an Address Conflict Notice (ACN) back in a certain timeout period. If, after 'n' number of retries, no ACN is received, the node may assume that address is not in use. This process is illustrated in Figure 2. But in networks where message delays cannot be bounded, use of timeouts can lead to unreliability. So duplicate addresses may occur in MANET. In case of merge, many nodes may have duplicate addresses thus overhead of the network would increase suddenly due to start of DAD process for every node. Address autoconfiguration method must treat it as a special case.

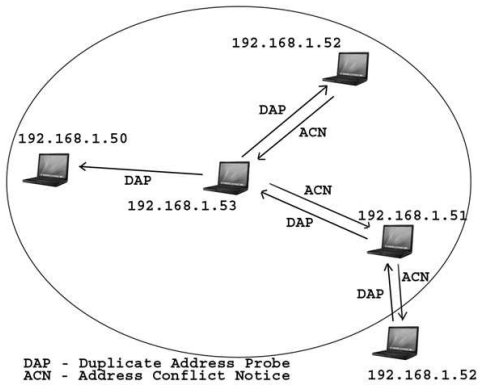


Figure 2: Duplicate Address Detection mechanism

[4] introduces StrongDAD & WeakDAD. StrongDAD allows at least one node to detect duplicate immediately after it has been chosen by another node. Practically it is not possible. WeakDAD is based on enhancement of link state routing. Each node of network owns a unique identifier. A node sends control packet indicating its

link state along with its identifier. Each node keeps state of the links it is connected to, corresponding addresses & identifiers. If a node N receives a control packet from a known address but with different identifier, then it has detected a duplicate. 'N' begins to announce duplicate and keep sending packets to the node it previously knows. MANETConf [7] proposes a reliable DAD process. It has two phases: initiation & validation. A new node (requester) takes help of a configured neighbor (initiator) to obtain address. Initiator broadcasts an address for the requester. All nodes have to answer this request. This ensures that requester would not use the address of a temporarily disconnected node. If a node does not answer after a number of tries, its address can be treated as unassigned.

2.4 Scalability of address auto-configuration

Number of messages need to be sent in order to assign address. These messages grow with number of network nodes leading to overhead. Overhead can also be due to merge and partition. Poor scalability can paralyze the network and lead to severe address leaks. In [8], a way to identify mobility patterns has been discussed. By using these mobility patterns, it ensures the reliability of any service in MANET. It includes a mechanism to allow servers in a MANET to detect the future partitions. It can enhance the guarantee of address autoconfiguration with minimal overhead, which leads to scaling of MANET on large number of nodes. Nevertheless these solutions use a strong centralized approach to detect partitions and their applicability as such may be questionable. It would be interesting to evaluate the possibility of making this partition detection in a distributed manner.The range of IP addresses should also be scalable. IP addresses should not run out of availability when a large number of nodes are joining at the same time e.g. during merging.

2.5 Secure and Authentic autoconfigurationn

Security & authenticity in mobile adhoc networks are hard to achieve due to frequently changing and fully decentralized system. Usage of security system can be dependant upon the application area. Possible address autoconfiguration attacks are[9]: In *Address Spoofing Attack*, a malicious node can freely choose any configured node as a victim, spoof its IP address & hijack its traffic. A *False Address Conflict Attack* may purposely transmit a false address conflict message to targeted victim. Since the victim cannot verify the authenticity of the purported address conflict, it may have to give up its current address and seek a new one. In *Address Exhaustion Attack*, an attacker could maliciously claim as many IP addresses as possible. If the attacker exhausts all valid IP addresses, a new node will not be able to get an address. In *Negative Reply Attack* an attacker may continuously send negative replies to prevent configuration of a new node. [9] uses self-authentication for secure autoconfiguration. By using one-way hash function, it binds a node's address with public key. Address owner can use corresponding public key to unilaterally authenticate itself. [10] employs the concept of challenge which obliges a node to answer a question to prove its identity. New node sends a request with its public key and a temporary identifier. Neighbors calculate a nonce that they return to new node, after having ciphered it with the public key. The mission of new node is then to return this nonce incremented to concerned nodes, after having ciphered it with its private key.

3 Research Trends

3.1 Electronic resources and search queries

The electronic resources used to find out trends are IEEE digital library and Springer's Lecture Notes in Computer Science. We found approximately 2700 articles on MANET, from the year 2002 to 2007. Out of these articles around 150 are related to address assignment. IEEE Xplore^{TM} [11] provides a powerful and convenient interface for searching. Papers can be searched using keywords, phrases or using boolean expression of various words. It has approximately 17,06,580 papers published in various IEEE / IET's periodicals, conference proceedings and books. Boolean expressions can be formulated using AND, OR and NOT operators. We used operators AND and OR. Keywords are selected as comprehensive as possible to represent the scope of each issue. In all these queries 'yyyy' varies from 2002 to 2007. Following type of queries are constructed:

- IEEE Xplore^{TM} query for searching papers on address assignment:
  (((((ip address)<in>metadata) < or > ((address assignment)< in >metadata) < or > ((auto configuration)< in >metadata)) < and > ((manet)< in >metadata))) < and > (pyr >= 'yyyy' <and> pyr <= 'yyyy')

- IEEE Xplore^{TM} queries for searching papers on various issues:
  (((((ip address)< in >metadata) < or > ((address assignment)< in >metadata) < or > ((auto configuration)< in >metadata)) < and > ((manet)< in >metadata) < and > ((partition)< in > metadata))) < and > (pyr >= 'yyyy < and > pyr <= 'yyyy')

  (((((ip address)< in >metadata) < or > ((address assignment)< in >metadata) < or > ((auto configuration)< in >metadata)) < and > ((manet)< in >metadata) < and > ((merging)< in > metadata))) < and > (pyr >= 'yyyy' < and > pyr <= 'yyyy')

Springer[12] is world's second-largest publisher of journals & books in the Science & Technology. Queries resulted in Springer journal articles as well as book chapters. Following are few queries:
Search For (Boolean) > ("mobile adhoc network") OR ("ad hoc network") Publication Date > Between Saturday, January 01, 2005 and Saturday, December 31, 2005
Search For (Boolean) > (("mobile adhoc network") OR ("adhoc network") OR ("ad hoc network")) AND (("address assignment") OR ("auto configuration") OR ("ip address")) Publication Date > Between Saturday, January 01, 2005 and Saturday, December 31, 2005
Search For (Boolean) > (("mobile adhoc network") OR ("ad hoc network")) AND (("address assignment") OR ("auto configuration") OR ("ip address")) AND ("partition")Publication Date > Between Sunday, January 01, 2006 and Sunday, December 31, 2006

3.2 Analysis Factors

Coverage of address issues in various papers is shown via statistical graphs. Papers related to autoconfiguration are rated on flexibility, reliability & heterogeneity. *Flexibility* is ease & speed with which auto-configuration methods can adapt to changing network conditions like address format etc. *Reliability* means autoconfiguration without any error and failure. *Heterogeneity* refers to dissimilar address formats and ways to assign them to nodes. It also refers to merging of dissimilar networks and enabling communication between them.

3.3 Results

Research trend statistics of address assignment versus MANET is shown in Figure 3: Left, in which number of papers published on address assignment is approximately 5.5 percent of the total MANET papers. Hence it can be concluded that address assignment issue still needs much more attention as compared to other MANET issues.

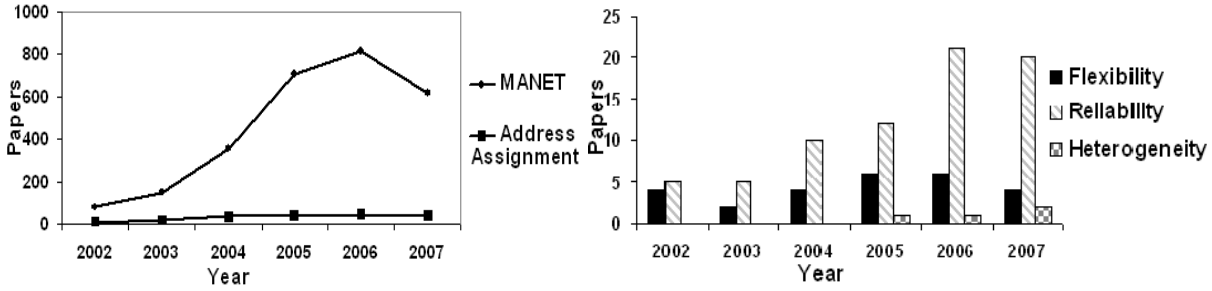


Figure 3: Left: MANET vs Address Assignment. Right: Various rating factors

Figure 4 gives us the quantity of papers covering various issues of address assignment and can be concluded that most of the time researchers are concerned with partitions of MANET, merging of several MANET and

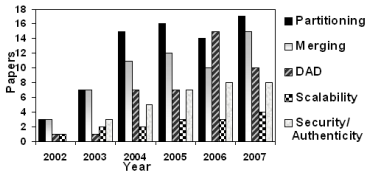


Figure 4: Trends in issues of address assignment

duplicate address detection. Very less attention has been paid to scalability and security / authenticity aspects of address assignment. So, there is requirement of research in these two aspects of address assignment.To rate the research trends we have chosen three factors: flexibility, reliability and heterogeneity. These factors can encourage researchers to formulate a new theory / explanation. If we look at the Figure 3: Right, then it is found that more attention is paid to reliability than flexibility & heterogeneity. There is need to develop the address assignment methods which can take care of heterogeneous networks.

4 Conclusion

Auto-configuration of IP address is a major issue that still needs attention from active researchers. To study the trends in this area two electronic resources viz IEEE and Springer's 'Lecture notes on Computer Science' are used. Approximately 150 papers are found to be covering the address assignment aspect of MANET. Problems such as partitioning, merging, duplicate address detection, scalability and security breaches have been recognized. Some of the proposed methods to solve these problems are available in literature. But still there is considerable requirement of effort to establish these methods. Particularly attention should be paid to scalability and secure / authentic address assignment. There is also requirement of developing these methods by taking into consideration heterogeneous environment of various MANET's.

References

[1] Harish Kumar & R.K Singla, "Issues in Mobile Ad-Hoc Networks," *Proceedings of the 41st Annual National Convention of Computer Society of India-CSI-2006*, November 23-25, 2006; pp 35-39

[2] Y. Sun & E.M. Belding-Royer, "A Study of Dynamic Addressing Techniques in Mobile Ad hoc Networks," *Wireless Comm. and Mobile Computing*, Vol 4, Issue 3, pp 315-329, John Wiley & Sons

[3] Matt W. Mutka, Lionel M. Ni & Hongbo Zhu, "Prophet Address Allocation for Large Scale MANETs," *Journal of Ad Hoc Networks*, Vol 4, No. 1, 2003, pp: 423-434.

[4] Nitin H. Vaidya, "Weak Duplicate Address Detection in Mobile Ad Hoc Networks," *Proceedings of 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing*, pp 206-216, 2002

[5] R. Droms, "A Dynamic host configuration protocol," *RFC 2131,* March 1997.

[6] Zhihua Hu & Baochun Li., "ZAL: Zero-Maintenance Address Allocation in mobile Wireless Ad Hoc Networks," *Proceedings of 25th ICDCS 2005*, pp 103-112, Columbus, Ohio, June 6-9, 2005

[7] S. Nesargi & R. Prakash,"MANETconf: Configuration of Hosts in a Mobile Ad Hoc Network," *INFOCOM 2002*, pp 1059-1068

[8] K.H. Wang & B. Li, "Group Mobility and Partition Prediction in Wireless Ad Hoc Networks," *IEEE International Conference on Communications (ICC)*, New York, NY, April 2002, pp 1017- 1021.

[9] P. Wang, D.S. Reeves & P. Ning, "Secure Address Auto-Configuration for Mobile Ad Hoc Networks," *Proceedings of 2nd Annual International Conference MobiQuitous 2005,* pp 519-522

[10] A. Cavalli & J.M. Orset, "Secure hosts Autoconfiguration in mobile ad hoc networks," *Proceedings of 24th International Conference on Distributed Computing Systems Workshop*,Vol 7, 2004, pp 809-814.

[11] *http:// www.ieee.org*

[12] *http://www.springerlink.com*

Harish Kumar, RK Singla, Siddharth Malhotra
Panjab University, Chandigarh, India
E-mail: harishk@pu.ac.in

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 358-363

Coloured Reconfigurable Nets For Code Mobility Modeling

Kahloul Laid, Chaoui Allaoua

Abstract: Code mobility technologies attract more and more developers and consumers. Numerous domains are concerned, many platforms are developed and interest applications are realized. However, developing good software products requires modeling, analyzing and proving steps. The choice of models and modeling languages is so critical on these steps. Formal tools are powerful in analyzing and proving steps. However, poorness of classical modeling language to model mobility requires proposition of new models. The objective of this paper is to provide a specific formalism "Coloured Reconfigurable Nets" and to show how this one seems to be adequate to model different kinds of code mobility.
Keywords: code mobility, modeling mobility, labeled reconfigurable nets, Coloured reconfigurable nets, mobile code design paradigms.

1 Introduction

Nowdays, code mobility is one of the attracting fields for computer science researchers. Code mobility technology seems an interest solution for distributed applications facing bandwidth problems, users' mobility, and fault tolerance requirement. Numerous platforms were developed [16]. Such platforms allow the broadcasting of this technology in many domains (information retrieving [8], e-commerce [10], network management [21], ...). Software engineering researches have provided some interest design paradigms influencing the development of the field. The most recognized paradigms [6] are: code on demand, remote evaluation, and mobile agent. To avoid ad-hoc development for code mobility software, many works attempt to propose methodologies and approaches ([15], [20], [13], ...). Indeed, these approaches are mostly informal. They lack in analyzing and proving system proprieties. Enhancing development process with formal tools was an attractive field in code mobility researches.

Traditional formal tools witch were massively used to model and analyze classical systems seem to be poor to deal with inherent proprieties in code mobility systems. Works on formal tools attempt to extended classical tools to deal with code mobility proprieties. The most important proposition can be found in processes algebra based model and state transition model. For the first one, π-calculus [12] is the famous one, and for the second, high-level Petri net (with many kinds) can be considered the good representative. π-calculus is an extension for CCS (communicating concurrent systems) [11]. CCS allows modeling a system composed of a set of communicating processes. This communication uses names (gates) to insure synchronization between processes. In $\pi-$calculus information can been exchanged through gates. The key idea is that this information can be also a gate. With this idea, processes can exchange gates. Once these gates received, they can be used by the receiver to communicate. In an extension of π-calculus, HOπ-calculus [14], processes can exchange other processes through gates (the exchanged processes called agents).

To model mobility with Petri nets, high level PNets were proposed. The most famous are Mobile Nets (variant of coloured Petri nets) [1] and Dynamic Petri nets. In mobile Petri nets, names of places can appear as tokens inside other places. Dynamic Petri nets extend mobile Petri nets. In this last one, firing a transition can cause the creation of a new subnet. With high-level Petri nets, mobility in a system is modeled through the dynamic structure of the net. A process appearing in a new environment is modeled through a new subnet created in the former net by firing a transition. Many extensions have been proposed to adapt mobile Petri net to specific mobile systems: Elementary Object Nets [17], reconfigurable nets [3], Nested Petri Nets [9], HyperPetriNets [2], ... With respect to [19], all these formalisms lack in security aspect specification. To handle this aspect in code mobility, recently Mobile Synchronous Petri Net (based on labeled coloured Petri net) are proposed [18].

The objective of this work is to present a new formalism based on Petri nets. Our formalism "Coloured reconfigurable nets" as an extension for our work "Labeled Reconfigurable Nets" [7]. We attempt to propose to model mobility in an intuitive and an explicit way. Mobility of code (a process or an agent) will be directly modeled through reconfiguration of the net. We allow adding and deleting of places, arcs, and transitions at run time. In this formalism, we introduce two kinds of specific transitions: calculi transitions and reconfigure transitions. A calculi transition takes as input a set of tokens (of type nets), it computes a set of places and arcs, and it outputs a set of tokens (of types: nets, places and arcs). The objective of this kind of transition is to prepare the reconfiguration of the net (migration of a net). A reconfigure transition takes as input tokens (of types: nets, places and arcs), it reconfigure the net by moving some subnets, places and arcs from one net towards another net. We propose that

these two kinds of transition allow modeling mobility in an explicit and more sophisticate manner. In this model we consider that nets, places or arcs can play as tokens. These tokens move from one place to another when some transitions are fired.

The rest of this paper is organized as follows. Section 2 starts by presenting the definition of the model "Coloured Reconfigurable Nets" or CRN. In section 3 we show how to model different design paradigms using CRN, for space reason, we consider only "mobile agent paradigms". In section 4, we conclude this work and give some perspectives, in section 5.

2 Definition of Coloured Reconfigurable Nets (CRN)

Coloured reconfigurable nets are an extension of labeled reconfigurable nets. Informally, a coloured reconfigurable net is a set of *environments* (blocs of *units*). Connections between these environments and their contents can be modified during runtime. A unit is a specific Petri net. A unit can contain three kinds of transitions (a unique *start transition*, a set of *ordinary transitions*, a set of *calculi transition* and a set of *reconfigure transitions.*, as shown in figures (1)-(4).

Figure 1: Start transition

Figure 2: Ordinary transition

Figure 3: Calculi transition

Figure 4: Reconfigure transition

Preconditions and post-conditions to fire a start or an ordinary transition are the same that in Petri nets. When a reconfigure transition is fired, a net N will be (re)moved from an environment E towards another environment E'. The net N, the environment E and E' are defined by a calculi transition witch must always precedes this one. After firing a reconfigure transition, the structure of the coloured reconfigurable net will be updated (i.e some places, arcs, and transitions will be deleted or added). Here after we give our formal definitions of the concepts: unit, environment and coloured reconfigurable net. After the definition, we present the dynamic aspect of this model.

To define coloured reconfigurable nets, we introduce firstly the definition of units and environment.

Definition 1. (Unit) A unit is a net U=(Σ, P, T, A, C, E) Σ : a finite set of types (colors); we denote by expr the set of expression that can be written using variables in sets of Σ. P: a finite set of places; T: a finite set of transitions. We have T=$\mathfrak{T} \cup \mathfrak{C} \cup \mathfrak{R}$. *Where* $\mathfrak{T}$: a set of ordinary transitions, $\mathfrak{T}=\{t_1, \ldots, t_n\}$. This set must contain a unique transition that we call a start transition. We denote this transition as strt, $\mathfrak{C}$: a set of calculi transitions, $\mathfrak{C}=\{c_1, \ldots, c_m\}$, $\mathfrak{R}$: a set of reconfigure transitions, $R=\{r_1, \ldots, r_p\}$. A: a set of arcs C: a color mapping from P to Σ. C joins to each place p a color c that we note C(p). E: an expression mapping from A to expr.

Definition 2. (Environment) an environment E is a quadruplet E=(GP, RP, U, A)

- GP = $\{gp_1, gp_2, \ldots, gp_s\}$ a finite set of specific places : "guest places";
- RP = $\{rp_1, rp_2, \ldots, rp_s\}$ a finite set of specific places : "resource places";
- U = { N_1, N_2, ... N_k} a set of nets. where T_1, T_2, ..., $T_k$are the sets of their transitions and StrT={$strt_1$, $strt_2$, ..., $strt_k$} is the set of their start transitions.
- A : a set of arcs, A$\subseteq$ GP x StrT$\cup$RPxT. Such that: T=T1$\cup$T2 $\cup \ldots \cup$Tk Remark : we say that a unit U is in an environment E iff the net U is a subnet of the net E.

Definition 3. (Coloured reconfigurable nets)
A coloured reconfigurable net (CRN) is couple N=(E, A), such that:
E: a finite set of environments;
A: a finite set (probably empty) of arcs; these arcs connect places (resp. transitions) from one environment to other transitions (resp. places) in another environment.

2.1 Dynamic of coloured reconfigurable nets:

To introduce the dynamic of CRN we consider three types (colors): $\mathfrak{P}$(set of places), $\mathfrak{N}$ (set of nets), and $\mathfrak{B}$ (set of arcs). We denote respectively by $\mathfrak{P}$*, $\mathfrak{N}$*, $\mathfrak{B}$* the three multisets of types $\mathfrak{P}$, $\mathfrak{N}$, $\mathfrak{B}$. We focus on the semantic of calculi and reconfigure transition.

Semantic of calculi transition:

A calculi transition must take as input three tokens of type N(two environments and one unit, the unit must be in one and only one of the two environments). Firing the calculi transition provides a token in the multi-sets $< \mathfrak{N}$*, $\mathfrak{P}$*, $\mathfrak{B}$*$>$. We can say that a calculi transition uses a set of nets to computes some arcs and places. At the output, it provides a composite token of the input nets and the computed arcs and places. In general, this token is used by a reconfigure transition.

If t is a calculi transition, and E_1, E_2, U are the input nets (U is in E_1), once t is fired it produce a token $<U+E_1+E_2, P, A>$ such that P and A are two multi-sets that can be defined like this: $P=\{p \in P_{E1} / p \notin P_U$ and $\exists\ t \in T_U$ such that $(p,t) \in A_{E1}$ or $(t,p) \in A_{E1}\ \}$, and

$A=\{a \in A_{E1} / a \notin A_U$ and $\exists\ (t,p) \in T_{E1} x P_U \cup T_U x P_{E1}$ and $a=(t,p)\}$.

Where P_N, A_N and T_N denote respectively places, arcs and transitions of a net N.

Semantic of reconfigure transition:

The objective of a reconfigure transition is to reconfigure the structure of the net. To be fired, a reconfigure transition takes as input a token in the multi-sets : $< \mathfrak{N}$*, $\mathfrak{N}$, $\mathfrak{P}$*, $\mathfrak{B}$*$>$. Firing a reconfigure transition will update the structure of the coloured reconfigurable nets that contains this transition in the following semantic:

If rt is a reconfigure transition and $<U+E_1, E_2, P, A>$ is an input token, to fire rt we impose that there exists a free place pg in GP_{E_2}; witch means: for each $t \in strT_{E_2}$, $(pg,t) \notin A_{E_2}$.

Once this condition is satisfied, firing rt changes N structurally such that:

If E_1 and E_2denote the same environment then N will be not changed;

Else:

1. The net U is removed from the net E1: $U_{E_2} \leftarrow U_{E_2} \cup \{U\}$;
2. The net U is added to the environment E_2: $U_{E_1} \leftarrow U_{E_1} / \{U\}$;
3. $A_{E_2} \leftarrow A_{E_2} \cup$(pg,strt); such that strt is the start transition for U.
4. Some elements of P are transformed from E_1 towards E_2, some other are cloned and some other will not be changed (resp for elements in A). These elements depend on the modeling case. In section 3, we show how these elements can be defined depending on the mobile code design paradigm to model.

3 Modeling Code Mobility with CRN

A mobile code system is composed of execution units (EUs), resources, and computational environments (CEs). EUs will be modeled as units and computational environments as environments. Modeling resources requires using a set of places.

Reconfigure transitions model mobility actions. The key in modeling mobility is to identify the unit to be moved, the target computational environment and the types of binding to resources and their locations. This information is supposed to be known before mobility. We use calculi transition as computing actions that compute this information. After computing these elements, the reconfigure transition updates the net by moving a unit from one environment to another. This moving must respect requirement for bindings to resources to insure the reliability of components on their new locations. Information concerning units, environments and bindings will be defined according to the resources types and to the three design paradigms: remote (REV) evaluation, code on demand (COD), and mobile agent (MA). For space reason, here after we show only model for mobile agent paradigm.

3.1 Mobile Agent

In mobile agent paradigm, execution units are autonomous agents. The agent itself triggers mobility. In this case, *rt* –the reconfigure transition- is contained in the unit modeling the agent.

Example 3.1: let E_1 and E_2 two computational environments. E_1 contains two agents, a mobile agent MA and a static agent SA_1; E_2 contains a unique static agent SA_2. The three agents execute infinite loops. MA executes actions $\{a_{11}, a_{12}, a_{13}\}$, SA_1 executes actions $\{a_{21}, a_{22}, a_{23}\}$, and SA_2 executes actions $\{a_{33}, a_{32}\}$. To be executed, a_{11} require a transferable resource TR_1 and a non-transferable resource bound by type PNR_1 witch is shared with a_{21}. a_{12} and a_{22} share a transferable resource bound by value, and a_{13} and a_{23} share a non-transferable resource NR_1. In E_1, SA_2 requires a non-transferable resource bound by type PNR_2 to execute a_{32}. PNR_2 has the same type of PNR_1.

The system will be modeled as a coloured reconfigurable net N. N contains two environments E_1, E_2 that model the two computational environments (CE_1 and CE_2). Units A_1, A_2 and A_3 will model MA, SA_1 and SA_2, respectively. In this case, the unit A_1 will contain a reconfigure transition *rt* and a calculi transition *ct*.

1. $E_1=(RP_1, GP_1, U_1, A_1)$; $RP_1=\{TR_1, PNR_1, VTR_1, NR_1\}$. $U_1=\{EU_1, EU_2\}$;
2. $E_2=(RP_2, GP_2, U_2, A_2)$; $RP_2=\{PNR_2\}$. $GP_2=\{PEU_1\}$.
3. *ct* will take as input tokens : E_1, A_1 and E_2. *ct* will provide the token: $<A_1+E_1, E_2, P, A>$. such that
   $P=TR_1+VTR_1$
   $A=(NR_1,a_{23})+(PNR_2,a_{21})$
4. *rt* takes as input $<A_1+E_1, E_2, P, A>$ and will remove A_1 and places in *P* from E_1 towards E_2. Arcs in *A* will be added to the N.

In the Figure 5, the types of places p^1, p^2, p^3 is N (set of nets). p^1 contains A_1, p^2 contains E_2 and p^3 contains E_2. The type of place p_{11} is $< \mathfrak{N}^*, \mathfrak{N}, \mathfrak{P}^*, \mathfrak{B}^* >$.

Figure 6 shows the model of this system after firing *rt*.

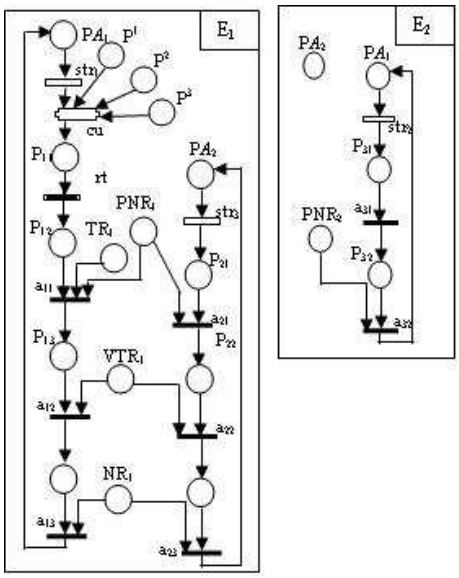


Figure 5: MA-model

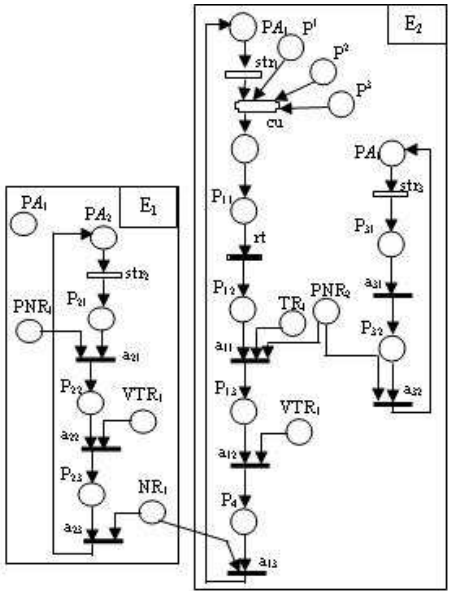


Figure 6: MA-Model after firing rt

4 Summary and Conclusions

Proposed initially to model concurrency and distributed systems, Petri nets attract searchers in mobility modeling domain. The ordinary formalism is so simple with a smart formal background, but it fails in modeling mobility aspects. Many extensions were proposed to treat mobility aspects. The key idea was to introduce mechanisms that allow reconfiguration of the model during runtime. Most works extend coloured Petri nets and borrow $\pi-$calculus or join calculus ideas to model mobility. The exchanging of names between processes in π-calculus is interpreted

as exchanging of place's names when some transitions are fired. This can model dynamic communication channels. In much formalism, mobility of processes is modeled by a net playing as token that moves when a transition is fired. All these mechanisms allow modeling mobility in an implicit way. We consider that the most adequate formalisms must model mobility explicitly. If a process is modeled as a subnet, mobility of this process must be modeled as a reconfiguration in the net that represents the environment of this process.

In this paper, we have presented a new formalism "Coloured reconfigurable nets". This formalism allows explicit modeling of computational environments and processes mobility between them. We have presented how this formalism allows, in a simple and an intuitive approach, modeling mobile code paradigms. We have focused on bindings to resources and how they will be updated after mobility. We believe that the present formalism is an adequate model for all kinds of code mobility systems. In our future works, we plan to focus on modeling and analyzing aspects. In modeling aspects, we are interested to handle problems such that modeling multi-hops mobility, process's states during travel, birth places and locations. On the analysis aspect, we are thinking about an encoding of our model in maude or mobile maude [5] in order an analysis automation of our models.

References

[1] Andrea Asperti and Nadia Busi. "Mobile Petri Nets". Technical Report UBLCS-96-10, Department of Computer Science University of Bologna, May 1996.

[2] M.A. Bednarczyk, L. Bernardinello, W. Pawlowski, and L. Pomello. "Modelling Mobility with Petri Hypernets". 17th Int. Conf. on Recent Trends in Algebraic Development Techniques, WADT'04. LNCS vol. 3423, Springer-Verlag, 2004.

[3] M. Buscemi and V. Sassone. "High-Level Petri Nets as Type Theories in the Join Calculus". In Proc. of Foundations of Software Science and Computation Structure (FoSSaCS '01), LNCS 2030, Springer-Verlag.

[4] Dianxiang Xu and Yi Deng, "Modeling Mobile Agent Systems with High Level Petri Nets". 0-7803-6583-6/00/ © 2000 IEEE.

[5] Francisco Dur n, Steven Eker, Patrick Lincoln and José Meseguer. "principles of mobile maude". In D.Kotz and F.Mattern, editors, Agent systems, mobile agents and applications, second international symposium on agent systems and applications and fourth international symposium on mobile agents, ASA/MA 2000 LNCS 1882, Springer Verlag. Sept 2000.

[6] Alfonso Fuggetta, Gian Pietro Picco and Giovanni Vigna, "Understanding Code Mobility". IEEE transactions on software engineering, vol. 24, no. 5, may 1998.

[7] Kahloul Laid and Chaoui Allaoua, "Labeled reconfigurable nets for modeling code mobility. In the proceeding of The International Arab Conference for Information technology (ACIT) 26-28/11/2007 in Syria.

[8] P. Knudsen, "Comparing Two Distributed Computing Paradigms, A Performance Case Study"; MS thesis, Univ. of Tromsϕ, 1995.

[9] I.A. Lomazova. "Nested Petri Nets"; Multi-level and Recursive Systems. Fundamenta Informaticae vol.47, pp.283-293. IOS Press, 2002.

[10] M. Merz and W. Lamersdorf, "Agents, Services, and Electronic Markets: How Do They Integrate?"; Proc. Int'l Conf. Distributed Platforms, IFIP/IEEE, 1996.

[11] R. Milner. "A Calculus of Communicating Systems". Number 92 in Lecture Notes in Computer Science. Springer Verlag, 1980.

[12] R. Milner, J. Parrow, and D. Walker. "A calculus of mobile processes". Information and Computation, 100:1–77, 1992.

[13] Reinhartz-Berger, I., Dori, D. and Katz, S. (2005) "Modelling code mobility and migration: an OPM/Web approach", Int. J. Web Engineering and Technology, Vol. 2, No. 1, pp.6–28.

[14] D. Sangiorgi and D. Walker. "The p-Calculus: A Theory of Mobile Processes". Cambridge University Press, 2001.

[15] Athie L. Self and Scott A. DeLoach. “Designing and Specifying Mobility within the Multiagent Systems Engineering methodology ” Special Track on Agents, Interactions, Mobility, and Systems (AIMS) at the 18th ACM Symposium on Applied Computing (SAC 2003). Melbourne, Florida, USA, 2003.

[16] Tommy Thorn, “Programming languages for mobile code”. Rapport de recherche INRIA, N ° 3134, Mars, 1997.

[17] R. Valk. “Petri Nets as Token Objects: An Introduction to Elementary Object Nets”. Applications and Theory of Petri Nets 1998, LNCS vol.1420, pp.1-25, Springer-Verlag, 1998.

[18] F. Rosa Velardo, O. Marroqn Alonso and D. Frutos Escrig. “Mobile Synchronizing Petri Nets: a choreographic approach for coordination in Ubiquitous Systems”. In 1st Int. Workshop on Methods and Tools for Coordinating Concurrent, Distributed and Mobile Systems, MTCoord’05. ENTCS, No 150.

[19] Fernando Rosa-Velardo. “Coding Mobile Synchronizing Petri Nets into Rewriting Logic”, this paper is electronically published in Electronic Notes in Theoretical Computer science URL: www.elsevier.nl/locate/entcs.

[20] Sutandiyo, W., Chhetri, M, B., Loke, S,W., and Krishnaswamy, S. “mGaia: Extending the Gaia Methodology to Model Mobile Agent Systems”, Accepted for publication as a poster in the Sixth International Conference on Enterprise Information Systems (ICEIS 2004), Porto, Portugal, April 14-17.

[21] D.J. Wetherall, J. Guttag, and D.L. Tennenhouse, “ANTS: A Toolkit for Building and Dynamically Deploying Network Protocols” Technical Report, MIT, 1997, in Proc. OPENARCH’98.

Kahloul Laid
Computer Science Department
Biskra University, Algeria
E-mail: kahloul2006@yahoo.fr

Chaoui Allaoua
LIRE Laboratory
Constantine University, Algeria

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 364-368

Computing Nash Equilibria by Means of Evolutionary Computation

Rodica Ioana Lung, Dan Dumitrescu

Abstract: The problem of detecting the Nash equilibria of a multi-player normal form game is solved by introducing a Nash based domination relation that enables evolutionary search operators to converge towards multiple solutions of a game.
Keywords: Nash equilibria, evolutionary computation

1 Introduction

A new evolutionary approach for solving normal form games is presented. The problem of detecting Nash equilibria (NE) is solved by using a new Nash based dominance concept. The proposed Nash domination concept in theoretically introduced. It is proved that solutions not dominated with respect to this relation coincide with the Nash equilibria of the game. Thus this concept enables the comparison of two solutions and allows the search of equilibria by means of evolutionary algorithms.

2 Prerequisites

Notations and basic notions related to game theory that are necessary for this work are presented in this section. A finite strategic game is defined by $\Gamma = ((N, S_i, u_i), i = 1, n)$ where:

- N represents the set of players, $N = \{1, ..., n\}$, n is the number of players;
- for each player $i \in N$, S_i represents the set of actions available to him, $S_i = \{s_{i1}, s_{i2}, ..., s_{im}\}$; $S = S_1 \times S_2 \times ... \times S_N$ is the set of all possible situations of the game;
- for each player $i \in N$, $u_i : S \to R$ represents the payoff function.

Let $U = \{u_1, ..., u_n\}$.
We denote by $(s_{i_j}, s^*_{-i})$the strategy profile obtained from s^* by replacing the strategy of player i with s_{i_j} i.e.

$$(s_{i_j}, s^*_{-i}) = (s^*_1, s^*_2, ..., s^*_{i-1}, s_{i_j}, s^*_{i+1}, ..., s^*_1).$$

2.1 Nash equilibrium

The most common concept of solution for a non cooperative game is the concept of Nash equilibrium [3, 5]. A collective strategy $s \in S$ for the game (S, U) represents a Nash equilibrium if no player has anything to gain by changing only his own strategy.

Definition 1. A strategy profile $s^* \in S$ is a Nash equilibrium (NE) if no deviation in strategy by any single player is profitable, that is, if for all players $i \in \{1, .., n\}$ and all strategies $s_{i_j} \in S_i$ the inequality

$$u_i(s_{i_j}, s^*_{-i}) \leq u_i(s^*),$$

holds.

Several methods to compute NE of a game have been developed. For a review on computing techniques for the NE see [3].

3 Nash dominance

A new concept of dominance for game strategies is defined in this section. This dominance concept is based on the Nash equilibrium concept.

Consider two strategy profiles s^* and s from S. We will introduce an operator $k : S \times S \rightarrow N$ that associates the pair (s^*, s) the cardinality of the set

$$\{i \in \{1,...,n\} | u_i(s_i, s^*_{-i}) \geq u_i(s^*), s_i \neq s^*_i\}.$$

This set is composed by the players i that would benefit if - given the strategy profile s^* - would change their strategy from s^*_i to s_i, i.e.

$$u_i(s_i, s^*_{-i}) \geq u_i(s^*).$$

Remark 2. It is obvious that for any $s^*, s \in S$, we have

$$0 \leq k(s^*, s) \leq n.$$

For any strategy profiles s^* and s there may by up to n players that can improve their payoffs by replacing their strategy s^*_i with s_i.

Definition 3. Let $x, y \in S$. We say the strategy profile x dominates the strategy profile y in Nash sense and we write $x \prec y$ if the inequality

$$k(x,y) < k(y,x),$$

holds.

Thus a strategy profile x dominates strategy profile y if there are less players that can increase their payoffs by switching their strategy from x_i to y_i then vice-versa. It can be said that strategy profile x is more stable (closer to equilibrium) then strategy y.

Remark 4. Two strategy profiles $x, y \in S$ can have the following relation:

1. either x dominates y, $x \prec y$ (we have $k(x,y) < k(y,x)$)
2. either y dominates x, $y \prec x$ (we have $k(x,y) > k(y,x)$)
3. or $k(x,y) = k(y,x)$ and x and y are considered indifferent (neither x dominates y nor y dominates x).

Definition 5. The strategy profile $s^* \in S$ is called non-dominated in Nash sense (NNS) if

$$\nexists s \in S, s \neq s^* \text{such that } s \prec s^*.$$

Definition 6. The set of all Nash nondominated strategy profiles is the set containing all nondominated strategies i.e.

$$NND = \{s \in S | s \text{ Nash non-dominated}\}$$

A characterization of NE using k operator is given in the next proposition.

Proposition 7. *A strategy profile $s^* \in S$ is a NE iff the equality*

$$\forall s \in S, k(s^*, s) = 0,$$

holds.

Proof. Let $s^* \in S$ be a NE. Suppose there exists a strategy profile $s \in S$ such that $k(s^*, s) = w$, $w \in \{1,...,n\}$. Therefore there exists $i \in \{1,...,n\}$ such that $u_i(s_i, s^*_{-i}) \geq u_i(s^*)$ and $s_i \neq s^*_i$, which contradicts the definition of NE.

For the second implication, let $s^* \in S$ be a strategy profile such that $\forall s \in S, k(s^*, s) = 0$. This means that for all $i \in \{1,...,n\}$ and for any strategy $s_{i_j} \in S_i$ we have $u_i(s_{i_j}, s^*_{-i}) \leq u_i(s^*)$. Based on definition 1 s^* is a NE. □

Proposition 8. *All NE are Nash nondominated solutions (NNS) i.e.*

$$NE \subseteq NNS.$$

Proof. Let $s^* \in S$ be a NE. Suppose that there exists a strategy profile $s \in S$ such that $s \prec s^*$. It follows that $k(s, s^*) < k(s^*, s)$. But $k(s^*, s) = 0$ (proposition 7) therefore we must have $k(s, s^*) < 0$ which is not possible since $k(s, s^*)$ denotes the cardinality of a set. □

Proposition 9. *All Nash nondominated solutions are NE, i.e.*

$$NNS \subseteq NE.$$

Proof. Let s^* be a nondominated strategy profile. Suppose s^* is not NE. Therefore there must exist (at least) one $i \in \{1,...,n\}$ and a strategy $s_{i_j} \in S_i$ such that

$$u_i(s_{i_j}, s^*_{-i}) > u_i(s^*),$$

holds. Let's denote by $q = (s_{i_j}, s^*_{-i})$. It means that $k(s^*, q) = 1$. But $k(q, s^*) = 0$. Therefore $k(q, s^*) < k(s^*, q)$ which means that $q \prec s^*$ thus the hypothesis that s^* is nondominated is contradicted. □

Using propositions 8 and 9 it is obvious that the next result holds:

Proposition 10. *The following relation holds:*

$$NE = NNS,$$

i.e. all NE are also Nash nondominated and also all Nash nondominated strategies are NE.

This result is very important as it allows us to locate NE by using evolutionary search operators based on the concept of Nash dominance.

4 Examples

A simple way to approach the problem is by using evolutionary algorithms designed for multiobjective optimization. The main reason for making this choice lyes in the similarities between a game in which players seek to maximize their playoffs and a multiobjective maximization problem.

Nondominated Sorting Genetic Algorithm II (NSGAII) is a state-of-art evolutionary algorithm for multiobjective optimization designed by Deb et al. [2]. By changing the domination testing procedure of this method (from testing Pareto domination to testing Nash domination) an evolutionary algorithm for computing NE called Nash NSGAII has been obtained. Another method used for testing the theory that adapting an evolutionary algorithm for multiobjective optimization using the Nash domination relation is Roaming Algorithm for Multiobjective Optimization (RAMO) [4]

4.1 Cournot model of duopoly

The Cournot model of duopoly [6] is considered. Let q_1 and q_2 denote the quantities of an homogeneous product, produced by two firms respectively. The market clearing price is $P(Q) = a - Q$, where $Q = q_1 + q_2$ is the aggregate quantity on the market. Hence we have

$$P(Q) = \begin{cases} a - Q, & \text{for } Q < a, \\ 0, & \text{for } Q \geq a. \end{cases}$$

We assume that the total cost for the firm i of producing quantity q_i is $C_i(q_i) = cq_i$. That is, there are no fixed costs and the marginal cost is constant at c, where we assume $c < a$. Suppose that the firms choose their quantities simultaneously. The payoff for the firm i is its profit, which can be written in the following form:

$$\begin{aligned} \pi_i(q_i, q_j) &= q_i P(Q) - C_i(q_i) \\ &= q_i \left[a - (q_i + q_j) - c \right]. \end{aligned}$$

The NE (q_1^*, q_2^*) of this game can be computed from the following relation:

$$\max_{0 \leq q_i \leq \infty} \pi_i(q_i, q_j^*) = \max_{0 \leq q_i \leq \infty} q_i \left[a - \left(q_i + q_j^* \right) - c \right].$$

The solution of this problem, i.e. the Nash equilibrium, is

$$q_1^* = q_2^* = \frac{a - c}{3}.$$

We considered the Cournot model from an optimization point of view considering the Nash domination. Thus the problem is to detect the Nash nondominated set of the problem:

$$\begin{cases} \pi_i(q_i,q_j) = q_i\left[a-(q_i+q_j)-c\right], \rightarrow \max \\ i = 1,2. \end{cases}$$

RAMO and NSGAII detected the Nash equilibrium of the Cournot duopoly model taking $a = 24$ and $c = 9$. This problem has one NE point

$$q = (5,5).$$

On 100 runs both techniques detected exactly this NE without any difficulty.

4.2 Partnership Game

Another model of game is the Partnership Game. There is a firm with two partners. The firm's profit depends on the effort each partner expends on the job and is given by

$$p = 4(x+y+cxy),$$

where x is the amount of effort expended by partner 1 and y is the amount of effort expended by partner 2. Assume that $x,y \in [0,4]$. The value $c \in [0,1/4]$ measures how complementary the tasks of the partners are. Partner 1 incurs a personal cost x^2 of expending effort, and partner 2 incurs cost y^2. Each partner selects the level of his effort independently of the other, and both do so simultaneously. Each partner seeks to maximize their share of the firm's profit (which is split equally) net of the cost of effort. That is, the payoff function for partner 1 is

$$u_1(x,y) = p/2 - x^2,$$

and that for partner 2 is

$$u_2(x,y) = p/2 - y^2.$$

The strategy spaces here are continuous $S_1 = S_2 = [0,4]$ and $S = [0,4] \times [0,4]$. This game has a unique Nash equilibrium for

$$x^* = 1/(1-c), y^* = 1/(1-c).$$

The two modified algorithms (NSGAII and RAMO) are used to solve this partnership game taking

$$c = \frac{1}{5}.$$

In this case the NE is

$$x^* = y^* = 5/4.$$

Both algorithms detected this NE without any difficulty.

4.3 Example - Game1

Consider the two player game [1] having the following payoff functions:

$$\begin{aligned} u_1(y_1,y_2) &= y_1 \\ u_2(y_1,y_2) &= (0.5-y_1)y_2 \end{aligned}$$

with $y_1 \in [0,0.5]$ and $y_2 \in [0,1]$. This game presents an infinity of NE of form $(0.5,\lambda)$, $\lambda \in [0,1]$.
Figure 1 presents results obtained after ten runs of NNSGAII with different random number generator seeds.

5 Conclusions and further work

A new domination relation based on the concept of Nash equilibrium is introduced. This relation can be incorporated in search tools such as nature inspired heuristics in order to guide their search towards the solutions of a normal form game. To the best of the authors knowledge, this is the first time a method of approaching NE by using a relation between strategy profiles has been proposed.

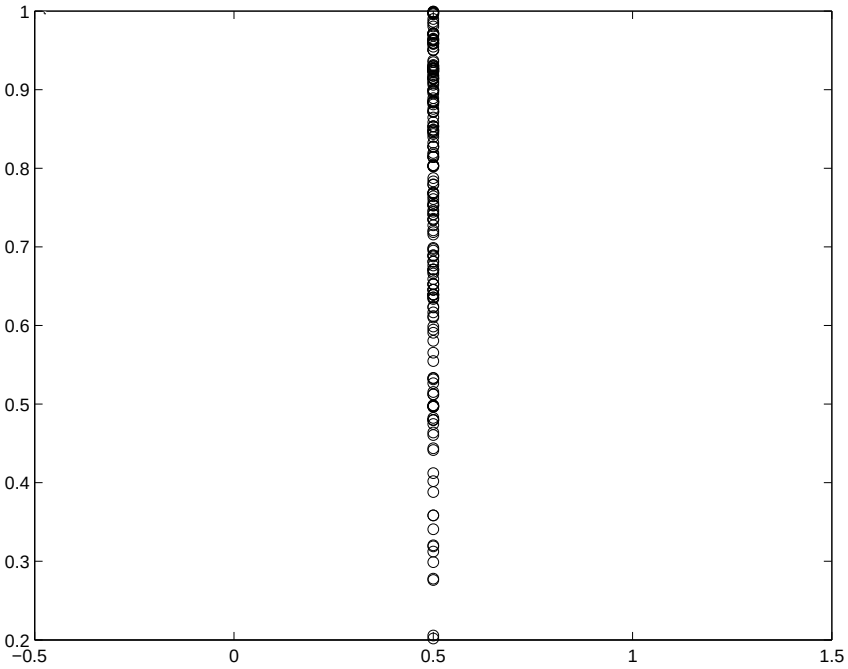


Figure 1: Example of results obtained using NNSGA-II for Game1

References

[1] Sophie Bade, Guillaume Haeringer and Ludovic Renou. More strategies, more Nash equilibria. Working Paper 2004-15, School of Economics University of Adelaide University, 5005 Australia, 2004.

[2] Kalyanmoy Deb, Samir Agrawal, Amrit Pratab, and T. Meyarivan. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel, editors, *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pages 849–858, Paris, France, 2000. Springer. Lecture Notes in Computer Science No. 1917.

[3] Richard D. McKelvey and Andrew McLennan. Computation of equilibria in finite games. In H. M. Amman, D. A. Kendrick, and J. Rust, editors, *Handbook of Computational Economics*, volume 1 of *Handbook of Computational Economics*, chapter 2, pages 87–142. Elsevier, 1 1996.

[4] Lung, R. I., Muresan, A. S., and Filip, D. A. Solving multi-objective optimization problems by means of natural computing with application in finance. In *Aplimat 2006* (Bratislava, February 2006), pp. 445–452.

[5] John F. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[6] www.wikipedia.com, last accessed May, 2006.

Rodica Ioana Lung
Babeş-Bolyai University
Faculty of Economics and Business Administration
58-60 T. Mihali Str.
Cluj-Napoca, Romania
E-mail: rlung@econ.ubbcluj.ro

D. Dumitrescu
Babeş-Bolyai University
Department of Computer Science
58-60 T. Mihali Str.
Cluj-Napoca, Romania
E-mail: ddumitr@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 369-374

Practices for Designing and Improving Data Extraction in a Virtual Data Warehouses Project

Ion Lungu, Manole Velicanu, Adela Bara, Vlad Diaconiţă, Iuliana Botha

Abstract: The problem of low performance in data extraction from data warehouse can be critical in the Business Intelligence projects because of the major impact in the using the data from data warehouse: if a BI report is taking a lot of time to run or the data displayed are no longer available for taking critical decisions, the project can be compromised. In this paper we present an overview of an implementation of a Business Intelligence project in a national company, the problems we confronted with and the techniques that we applied to reduce the cost of execution for improving query performance in this decisional support system. We'll present several techniques that we applied to reduce queries' execution time and to improve the performance of the BI analyses and reports.
Keywords: Business Intelligence, Virtual data warehouse, Data extraction, Query optimization, Partitioning techniques, Indexes, Analytical functions

1 Introduction

Business Intelligence Systems allows users to manipulate large sets of data in real time manner in order to analyze and take decisions. So, the main goal of Business Intelligence Systems (BIS) is to assist managers, at different levels in the organization, in taking decisions and to provide in real time representative information, to help and support them in their activities such as analyzing departmental data, planning and forecasting activities for their decision area [1]. In essence, managers at every departmental level can have a customized view that extracts information from transactional sources and summarizes it into meaningful indicators. These systems usually work with large sets of data and require a short response time and if you consider using analytical tools like OLAP against virtual data warehouses then you have to build your system through SQL queries and retrieve data directly from OLTP systems. In this case, the large amount of data in ERP systems may lead to an increase of responding time for BIS. That's why you should consider applying optimization techniques in order to improve the BI system's performance.

2 Problems in developing a BI project involving virtual data warehouses

In a research project that we conducted in one of the multinational companies from our country we applied the concepts mentioned in this paper and also tried to meet the requests from the executives and managers of the company. For the project life cycle we applied the framework described in the book Executive Information Systems [2], tailored for the specific needs of our project. With BI techniques, like data warehouse, OLAP, data mining, portal we succeeded to implement the BI system's prototype and to validate it with the managers and executives. The BIS gathers data, using the ERP system partially implemented in the organization to extract data from different functional areas or modules such as: financial, inventory, purchase, order management or production. Information from these functional modules within the ERP system is managed by a relational database management system - Oracle Database Enterprise Edition 10g. From the relevant data sources we applied an ETL (extract, transform and load) process to transport data from source to destination. In this step we had to choose between the two data warehouses solutions: stored data and virtual extraction. After a comparative analysis between these techniques we choose the second solution. The major elements in this choice were that the ERP system was not yet fully implemented and there are many more changes to do, the amount of data is not so large - 3 millions records from January 2007 to August 2007, and implementing a virtual data warehouse is fastest and needs a lower budget then a traditional data warehouse. We are also considering the development of a traditional data warehouse after testing and implemented the actual prototype. So, we build the virtual data warehouse based on a set of views that collects data from the ERP system based on an ETL process that we designed. After we developed the analytical decisional reports and test them in a real organizational environment with over 100 users, we measured the performance of the system and the main problem was the high cost of execution. These analytical reports consumed over 80 percent of the total resources allocated for the ERP and BI systems. Also, the critical moment when the system was breaking down was at the end of each month when all transactions from functional modules were posted to the General

Ledger module. Testing all parameters and factors we concluded that the major problem was in the data extraction from the views. First solution was to rewrite the views and build materialized views and semi-aggregate tables. Data sources are loaded in these tables by the ETL process periodically, at the end of the month after posting to the General Ledger or at user request. The ETL process contains a set of procedures grouped in three types of packages: extraction, transforming and loading. The target data are stored in tables used directly by the analytical reports. A benefit of this solution is that it eliminates the multiple joins from the views and also that we can use the ETL process to load data in future data warehouse. After these operations were completed we re-tested the systems under real conditions. The time for data extraction was again too long and the costs of executions consumed over 50 percent of total resources. Next we considered using some optimization techniques like: table partitioning, indexing, using hints and using analytical functions instead of data aggregation in some reports. In the following section we describe these techniques and provide a comparative analysis of some of our testing results.

3 Applying optimization techniques

3.1 Partitioning techniques

The main objective of the partitioning technique is to radically decrease the amount of disk activity and to limit the amount of data to be examined or operated on and to enable parallel execution required to perform Business Intelligence queries against virtual data warehouses. Tables are partitioning using a partitioning key that is a set of columns which will determine by their conditions in which partition a given row will be store. Oracle Database 10g on which the ERP is implemented provides three techniques for partitioning tables:

Range Partitioning - specify by a range of values of the partitioning key;

List Partitioning - specify by a list of values of the partitioning key;

Hash Partitioning - a hash algorithm is applied to the partitioning key to determine the partition for a given row;

Also, there can be use sub partitioning techniques in which the table in first partitioned by range/list/hash and then each partition is divided in sub partitions:

Composite Range-Hash Partitioning - a combination of Range and Hash partitioning techniques, in which a table is first range-partitioned, and then each individual range-partition is further sub-partitioned using the hash partitioning technique;

Composite Range-List Partitioning - a combination of Range and List partitioning techniques, in which a table is first range-partitioned, and then each individual range-partition is further sub-partitioned using the list partitioning technique;

Index-organized tables can be partitioned by range, list, or hash. [5] In our case we consider evaluating each type of partitioning technique and choose the best method that can improve the BI system's performance.

Thus, we create two tables based on the main table which is used by some analytical reports and compare the execution cost obtained by applying the same query on them. First table BALANCE_RESULTS_A contained non partitioned data and is the target table for an ETL sub-process. It counts 55000 rows and the structure is shown below in the scripts. The second table BALANCE_RESULTS_B is a range partitioned table by column ACC_DATE which refers to the accounting date of the transaction. This table has four partitions as you can observe from the script below. Then, we create a third table which is partitioned and that contained also for each range partition four list partitions on the column "Division" which is very much used in data aggregation in our analytical reports.

TABLE:	table_A	table_B		table_C		
	Not partitioned	Partition range by date on column "ACC_DATE"		Partition range by date with four list partions on column "DIVISION"		
QUERRY:		Without partition clause	Partition (QT1)	Without partition clause	Partition (QT1)	Sub-partition (QT1_AFO)
Select * from ...	100	121	-	224	-	-
where extract (**month from** acc_date) =1;	103	124	42	227	72	72
... **and** division='h.AFO divizii'	101	122	42	10	5	72
select sum(acc_d) TD, **sum**(acc_c) TC **from** balance_results_a a **where** extract (**month from** acc_date) =1 **and** division='h.AFO divizii'	101	122	42	10	5	72
... **and** management_unit ='MTN'	101	122	42	225	72	72
select /*+ *USE_HASH (a u)*/ a.*, u.**location**,u.country, u.region **from** balance_results_a a ,management_units u **where** a.management_unit=u.managem ent_unit **and** extract (**month from** acc_date) =1	106	127	46	231	76	76
... and a.division = 'h.AFO divizii'	105	126	45	14	9	75
.../*+ *USE_NL (a u)*/	191	212	71	100	8	101
.../*+ *USE_NL (a u)*/ ...*WITH INDEXES*	151	172	58	60	6	88
.../*+ *USE_MERGE (a u)*/ ...*WITH INDEXES*	104	125	45	13	8	75
...**and** u.management_unit ='MTN'	104	125	45	75	9	75
.../*+ *USE_NL (a u)*/	104	125	45	11	8	75
.../*+ *USE_NL (a u)*/ ...*WITH INDEXES*	102	123	43	11	6	73
.../*+ *USE_MERGE (a u)*/ ...*WITH INDEXES*	102	123	43	11	6	73

Figure 1: Comparative analysis results - the grey marked ones have the best execution cost of the current query

CodeSection 1. create table balance_results_b (ACC_DATE date not null, PERIOD varchar2(15) not null, ACC_D number, ACC_C number, ACCOUNT varchar2(25), DIVISION varchar2(50), SECTOR varchar2(100), MANAGEMENT_UNIT varchar2(100)) partition by range (ACC_DATE) (partition QT1 values less than (to_date('01-APR-2007', 'dd-mon-yyyy')), partition QT2 values less than (to_date('01-JUL-2007', 'dd-mon-yyyy')), partition QT3 values less than (to_date('01-OCT-2007', 'dd-mon-yyyy')), partition QT4 values less than (to_date('01-JAN-2008', 'dd-mon-yyyy')));

create table balance_results_C (ACC_DATE date not null, PERIOD varchar2(15) not null, ACC_D number, ACC_C number, ACCOUNT varchar2(25), DIVISION varchar2(50), SECTOR varchar2(100), MANAGEMENT_UNIT varchar2(100)) partition by range (ACC_DATE) subpartition by list (DIVISION) (partition QT1 values less than (to_date('01-APR-2007', 'dd-mon-yyyy')) (subpartition QT1_OP values ('a.MTN', 'b.CTM', 'c.TRS', 'd.WOD', 'e.DMA'), subpartition QT1_GA values ('f.GA op', 'g.GA corp'), subpartition QT1_AFO values ('h.AFO div', 'i.AFO corp'), subpartition QT1_EXT values ('j.EXT', 'k.Imp')), partition QT2 values less than (to_date('01-JUL-2007', 'dd-mon-yyyy')) (subpartition QT2_OP values ('a.MTN', 'b.CTM', 'c.TRS', 'd.WOD', 'e.DMA'), subpartition QT2_GA values ('f.GA op', 'g.GA corp'), subpartition QT2_AFO values ('h.AFO div', 'i.AFO corp'), subpartition QT2_EXT values ('j.EXT', 'k.Imp')), partition QT3 values less than (to_date('01-OCT-2007', 'dd-mon-yyyy')) (subpartition QT3_OP values ('a.MTN', 'b.CTM', 'c.TRS', 'd.WOD', 'e.DMA'), subpartition QT3_GA values ('f.GA op', 'g.GA corp'), subpartition QT3_AFO values ('h.AFO div', 'i.AFO corp'), subpartition QT3_EXT values ('j.EXT', 'k.Imp')), partition QT4 values less than (to_date('01-JAN-2008', 'dd-mon-yyyy')) (subpartition QT4_OP values ('a.MTN', 'b.CTM', 'c.TRS','d.WOD', 'e.DMA'), subpartition QT4_GA values ('f.GA op', 'g.GA corp'), subpartition QT4_AFO values ('h.AFO div', 'i.AFO corp'), Subpartition QT4_EXT values ('j.EXT', 'k.Imp')));

Analyzing the decision support reports we choose a sub-set of queries that are always performed and which are relevant for testing the optimization techniques. We run these queries on each test table: A, B and C and compare the results in figure 1.

In conclusion, the best technique in our case is to use table C instead table A or table B, that means that partitioning by range of ACC_DATE with type DATE and then partitioning by list of DIVISION with type VARCHAR2

is the most efficient method. Also, we obtained better results with table B partitioned by range of ACC_DATE than table A non-partitioned.

3.2 Indexing and applying hints

Oracle uses indexes to avoid the need for large-table, full-table scans and disk sorts, which are required when the SQL optimizer cannot find an efficient way to service the SQL query.

The oldest and most popular type of Oracle indexing is a standard b-tree index, which excels at servicing simple queries. The b-tree index was introduced in the earliest releases of Oracle and remains widely used with Oracle. While b-tree indexes are great for simple queries, they are not very good for the following situations:

Low-cardinality columns with less than 200 distinct values do not have the selectivity required in order to benefit from standard b-tree index structures.

No support for SQL functions. The B-tree indexes are not able to support SQL queries using Oracle's built-in functions. Oracle 9i provides a variety of built-in functions that allow SQL statements to query on a piece of an indexed column or on any one of a number of transformations against the indexed column.

Oracle bitmap indexes are very different from standard b-tree indexes. In bitmap structures, a two-dimensional array is created with one column for every row in the table being indexed. Each column represents a distinct value within the bit mapped index. This two-dimensional array represents each value within the index multiplied by the number of rows in the table. At row retrieval time, Oracle decompresses the bitmap into the RAM data buffers so it can be rapidly scanned for matching values. These matching values are delivered to Oracle in the form of a Row-ID list, and these Row-ID values may directly access the required information. The real benefit of bit mapped indexing occurs when one table includes multiple bit mapped indexes. Each individual column may have low cardinality. The creation of multiple bit mapped indexes provides a very powerful method for rapidly answering difficult SQL queries.

One of the most important advances in Oracle indexing is the introduction of function-based indexing. Function-based indexes allow creation of indexes on expressions, internal functions, and user-written functions in PL/SQL and Java. Function-based indexes ensure that the Oracle designer is able to use an index for its query.

Oracle indexes can greatly improve query performance but there are some important indexing concepts to review: index clustering and index block sizes. Indexes that experience lots of index range scans of index fast full scans (as evidence by multi block reads) will greatly benefit from residing in a 32 k block size. Today, most Oracle tuning experts utilize the multiple block size feature of Oracle because it provides buffer segregation and the ability to place objects with the most appropriate block size to reduce buffer waste [6].

The optimizer decision to perform a full-table vs. an index range scan is influenced by the clustering factor (located inside the dba_indexes view), db_block_size, and avg_row_len. It is important to understand how the optimizer uses these statistics to determine the fastest way to deliver the desired rows. Conversely, a high clustering_factor, where the value approaches the number of rows in the table (num_rows), indicates that the rows are not in the same sequence as the index, and additional I/O will be required for index range scans. As the clustering factor approaches the number of rows in the table, the rows are out of sync with the index.

When a SQL statement is executed the query optimizer determines the most efficient execution plan after considering many factors related to the objects referenced and the conditions specified in the query. The optimizer estimates the cost of each potential execution plan based on the statistics available in the data dictionary for the data distribution and storage characteristics of the tables, indexes, and partitions accessed by the statement and it evaluates the execution cost. This is an estimated value depending on resources used to execute the statement which includes I/O, CPU, and memory [4]. This evaluation is an important factor in the processing of any SQL statement and can greatly affect execution time.

We can override the execution plan of the query optimizer with hints inserted in SQL statement. A SQL statement can be executed in many different ways, such as *full table scans, index scans, nested loops, hash joins and sort merge joins*. We can set the parameters for query optimizer mode depending on our goal. For BIS, time is one of the most important factor and we should optimize a statement with the goal of best response time. To set up the goal of the query optimizer we can use one of the hints that can override the OPTIMIZER_MODE initialization parameter for a particular SQL statement [5]. The optimizer first determines whether joining two or more tables having UNIQUE and PRIMARY KEY constraints and places these tables first in the join order. The optimizer then optimizes the join of the remaining set of tables and determinate the cost of a join depending on the following methods:

Hash joins are used for joining large data sets and the tables are related with an equality condition join. The optimizer uses the smaller of two tables or data sources to build a hash table on the join key in memory and then

it scans the larger table to find the joined rows. This method is best used when the smaller table fits in available memory. The cost is then limited to a single read pass over the data for the two tables.

Nested loop joins are useful when small subsets of data are being joined and if the join condition is an efficient way of accessing the second table.

Sort merge joins can be used to join rows from two independent sources. Sort merge joins can perform better than hash joins if the row sources are sorted already and a sort operation does not have to be done.

We compare these techniques using hints in SELECT clause and based on the results in table 1 we conclude that the Sort merge join is the most efficient method when table are indexed on the join column for each type of table: non-partitioned, partitioned by range and partitioned by range and sub partitioned by list. The significant improvement is in sub partitioned table in which the cost of execution was drastically reduce at only 6 points compared to 102 points of non-partitioned table. Without indexes, the most efficient method is hash join with best results in partitioned table and sub partitioned table.

3.3 Re-write aggregate queries in an analytical mode

Aggregate functions applied on a set of records return a single result row based on groups of rows. Aggregate functions such as SUM, AVG and COUNT can appear in SELECT statement and they are commonly used with the GROUP BY clauses. In this case the set of records is divided into groups, specified in the GROUP BY clause. Aggregate functions are used in analytic reports to divide data in groups and analyze these groups separately and for building subtotals or totals based on groups.

Analytic functions process data based on a group of records but they differ from aggregate functions in that they return multiple rows for each group. In the latest versions in addition to aggregate functions Oracle implemented analytical functions to help developers building decision support reports [5]. The group of rows is called a window and is defined by the analytic clause. For each row, a sliding window of rows is defined and it determines the range of rows used to process the current row. Window sizes can be based on either a physical number of rows or a logical interval, based on conditions over values [3]

Analytic functions are performed after completing operations such joins, WHERE, GROUP BY and HAVING clauses, but before ORDER BY clause. Therefore, analytic functions can appear only in the select list or ORDER BY clause [4].

Analytic functions are commonly used to compute cumulative, moving and reporting aggregates. The need for these analytical functions is to provide the power of comparative analyses in the BI reports and to avoid using too much aggregate data from the virtual data warehouse.

Thus, we can apply these functions to write simple queries without grouping data like the following example in which we can compare the amount of current account with the average for three consecutive months in the same division, sector and management unit, back and forward:

CodeSection 2. select period, division, sector, management_unit, acc_d, avg(acc_d) over (partition by division, sector, management_unit order by extract (month from acc_date) range between 3 preceding and 3 following) avg_neighbors from balance_results_a

Or we can obtain the previous three months cumulative accountings applying analytical sum function:

CodeSection 3. select period, division, sector, acc_d, sum(acc_d) over (partition by division, sector order by extract (month from acc_date) range between 3 preceding and current row)d3_months_preceding, acc_c, sum(acc_c) over (partition by division, sector order by extract (month from acc_date) range between 3 preceding and current row)c3_months_preceding from balance_results_a where extract (month from acc_date) =6

Analyzing the execution cost obtained in each sample table (A, B and C) we can observe that the lowest cost is o table A (104) followed by B (125) and the highest cost in on table C (225). So, the conclusion is that it is not a very good solution to partition tables when applying analytical functions.

4 Summary and Conclusions

The virtual data warehouse that we designed and tested is based on a set of views that extracts, joins and aggregates rows from many tables from an ERP system. In order to develop this decisional system we have to build analytical reports based on SQL queries. But as the data extraction is the major time and cost consuming job we had to search for a solution. So, we tested several techniques that improved the performance. In this paper we presented some of these techniques, like table partitioning, using hints, loading data from the online views

in materialized views or in tables in order to reduce multiple joins and minimized the execution costs. Also, for developing reports an easy and important option is to choose analytic functions for predictions (LAG and LEAD), subtotals over current period (SUM and COUNT), classifications or ratings (MIN, MAX, RANK, FIRST_VALUE and LAST_VALUE). Another issue that is discussed in this article is the difference between data warehouses in which the data are stored in aggregate levels and virtual data warehouse. Our conclusion is that classical warehouse provides performance as virtual warehouse provides ease in development and keeps the costs down. Even if the purpose of a project is developing a classical warehouse, it's better to create a virtual warehouse before doing that.

References

[1] Lungu Ion, Bara Adela, Fodor Anca, "Business Intelligence tools for building the Executive Information Systems," *5thRoEduNet International Conference, Lucian Blaga University*, Sibiu, June, 2006.

[2] Lungu Ion, Bara Adela, *Executive Information Systems,* ASE Publisher, 2007.

[3] Lungu Ion, Bara Adela, Diaconita Vlad, "Building Analytic Reports for Decision Support Systems - Aggregate versus Analytic Functions," *Economy Informatics Review*, nr.1-4, pg. 17-20, ISSN 1582-7941, 2006.

[4] Oracle Corporation, "Database Performance Tuning Guide 10g Release 2 (10.2)" *Oracle Publications*, Part Number B14211-01, 2005.

[5] Oracle Corporation, "Oracle Magazine" *Oracle Publications*, 2006.

[6] Donald K. Burleson, *Oracle Tuning,* ISBN 0-9744486-2-1, 2006.

[7] Oracle Corporation Documentation, *www.oracle.coms*

Ion Lungu, Manole Velicanu, Adela Bara, Vlad Diaconiţă, Iuliana Botha
Academy of Economic Studies
Economic Informatics Department
Bucharest, Romania
E-mail: ion.lungu@ie.ase.ro, manole.velicanu@ie.ase.ro, bara.adela@ie.ase.ro
diaconita.vlad@csie.ase.ro iuliana.botha@ie.ase.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 375-380

CRM Kernel-based Integrated Information System for a SME: An Object-oriented Design

Vasile Lupşe, Ioan Dziţac, Simona Dziţac, Adriana Manolescu, Mişu-Jan Manolescu

Abstract: We propose an object-oriented design of an information integrated system for a SME. Our design is based on a kernel which implements CRM functions. This kernel is conceived as an independent subsystem and it is the first to be implemented. The others added subsystems are designed in a way that they will gravitate around the kernel. This type of integrated information system is developed in the iterative and incremental steps. CRM kernel implements basic functionality of the system, which stresses the financial partner relationships management.
Keywords: Customer Relationship Management (CRM), Small and/or Medium Enterprise (SME), Object-oriented Design (OOD).

1 Introduction

Competition challenge and economic progress of a company is strongly determinate by the manager's view regarding the implementation of information integrated systems for managerial decision-making and computer-assisted production process. The specific matters of the impact of the information society over a SME are presented in [1]-[4].

In the design process of an information system through the traditional method, each subsystem of an information system is independently developed. The subsystems are conceived in such a manner as to be distinct applications, implemented in different moments, trough different soft paradigms and tools. The integration of data by these applications is a bottom-up process that needs a common communication support between the subsystems (traditionally obtained by documents on paper).

The difficulties of such an approach are generated, firstly, by the absence of a common data model. The communication between the subsystems based on classical documents produce a data redundancy, accompanied by inconsistency premises.

Starting with these reasons, in this paper we propose another approach in designing an integrated information system for a wholesale trade small and medium enterprise. This approach is based on a CRM kernel [1]. This kernel is conceived as an independent subsystem. This type of integrated information system is developed in some of the iterative and incremental steps. CRM kernel is the first to be implemented. The other added subsystems are designed in a way that they will gravitate around the kernel.

2 The architecture of a CRM kernel based system

The architecture of a CRM kernel based system is presented in figure 2 and it has the following characteristics:

a) The central element of the system (both strictly speaking and figurative) is the CRM kernel, which stresses the relationship with the partners (customers and providers, especially);

b) Some traditional subsystems of the enterprise are left out of the new architecture (the *Commercial* subsystem, for example).

c) The new subsystems reflect better the marketing aspects of an enterprise (thus, the Commercial subsystem functions are now distributed in the three new subsystems: Sales, Supply and Marketing).

d) The Management subsystem, trough the integrated system facilities, offers to the manager's operative and synthetic information (for decision-making: operative, tactic, strategic).

Therefore, the Management subsystem has its own input data (contacts, address book, meetings, events), that are included in the general data model.

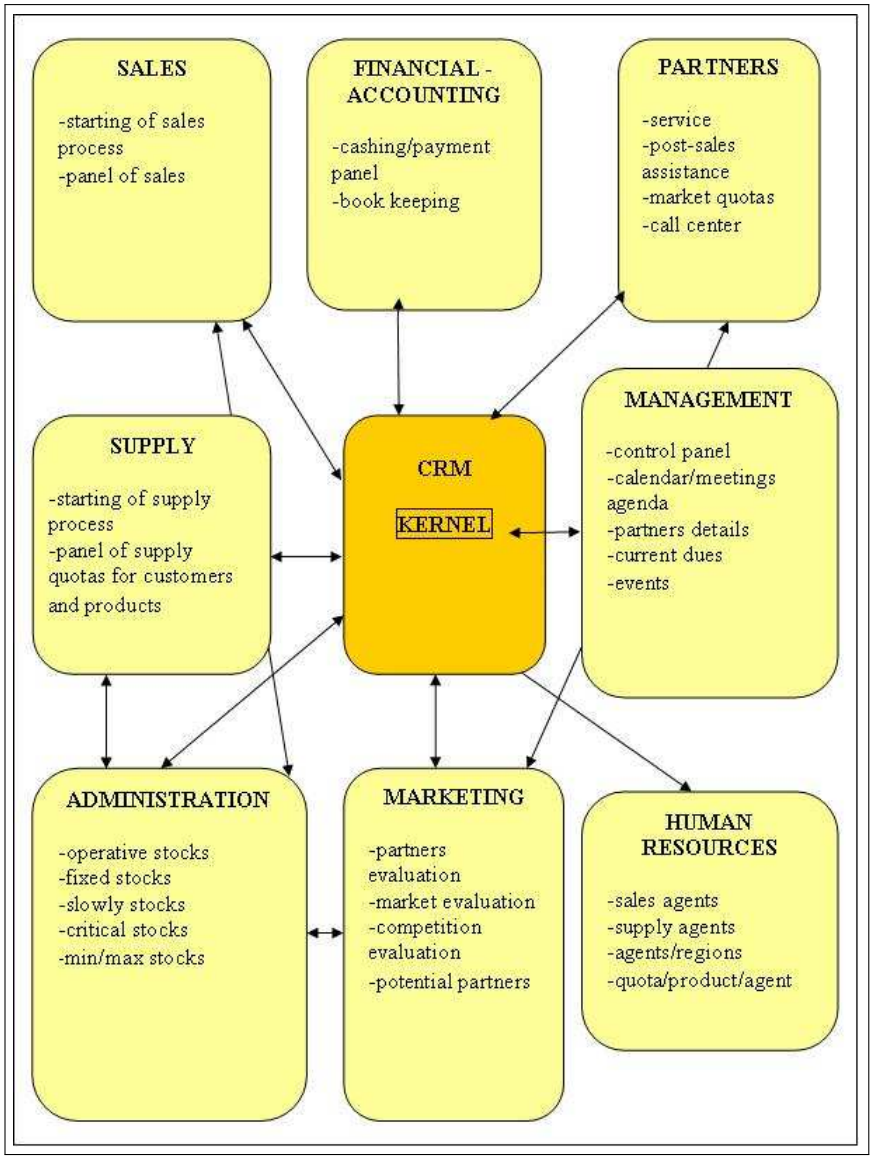


Figure 1: Architecture of a CRM kernel-based integrated information system for an en-gross commercial company

3 Object -oriented analyze for CRM kernel

After analyzing the previous requests, we will deduce the general used cases and the static model classes. The classes' names are nouns from the problem's domain vocabulary.

Table 1 includes the identified classes and attributes. We are mentioning that in this development phase, only the classes from the domain's problem are analyzed. The static model of the CRM kernel is presented in figure 2.

3.1 Data base paradigm

The data base scheme contains, as it is known, the explicit definitions of the modeling entities from the data base together with their attributes and the relationships between the entities. Therefore, the scheme includes the afferent tables of the synoptic view of data base, together with their links.

The transition from the class diagram to the data base scheme (figure 3) is natural, because the classes are normalized. Thus, they are mapping naturally on the related data base tables, named DB-CRM. In this process, the class names conversion (nouns at singular) in table names (nouns at plural) is trivial.

Table 1: CRM Kernel: Primary using cases

Symbol use case	Name
UC1	Issuing invoice to the customer
UC2	Incoming invoice from the customer
UC3	Receiving invoice from the supplier
UC4	Payment invoice to the supplier
UC5	Generating reports
UC6	Maintaining lists

Table 2: CRM Kernel: Classes and attributes from the static model -analyze

Class name	Attributes
Partner	Fiscal code, name, town, street and number, county, country, zip code, bank, account, phone, fax, email, web URL
Issued invoice	series and number, date, partner (customer), bank account, incoming sum, settling day
Received invoice	series and number, date, partner (supplier), bank account, payment sum, settling day
Incoming	partner (customer), issued invoice, incoming sum, type of incoming document, series and number of the document, date of incoming
Payment	partner (supplier), received invoice, paid sum, type of payment document, series and number of the document, date of payment

3.2 Extended functionalities by iterative and incremental method

Due to the dimension of this paper, this section presents only one CRM kernel extension discussed in the previous section, by analyzing only one subsystem presented in figure 2 (Partners).

The extension is considered the new way that integrates with the existing ones and uses the common data base. For this extension (and for other extensions), the data base will be enriched with new tables, which contain data and specific relations of this (those) extensions.

It is good to specify, from the beginning, that the partner could be a customer, a supplier or competitor for the enterprise.

The Partners subsystem was conceived to permit the adding at the existing functionality of the following specific functions:

A) Administration of the exchanged messages with the partners. Inside this function there were included opera-

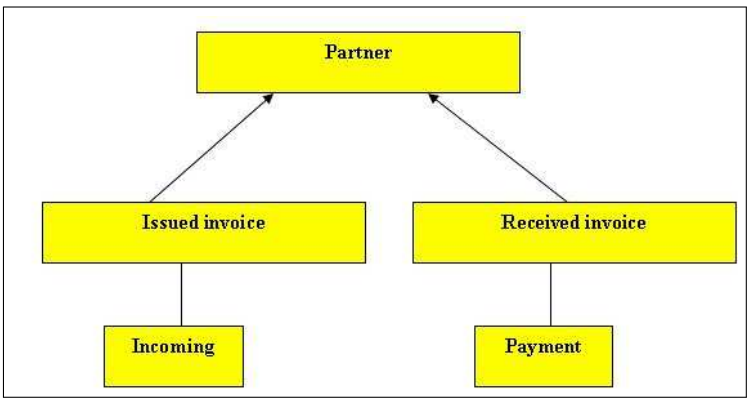


Figure 2: Class diagram of the CRM kernel - specific classes of the problem's domain

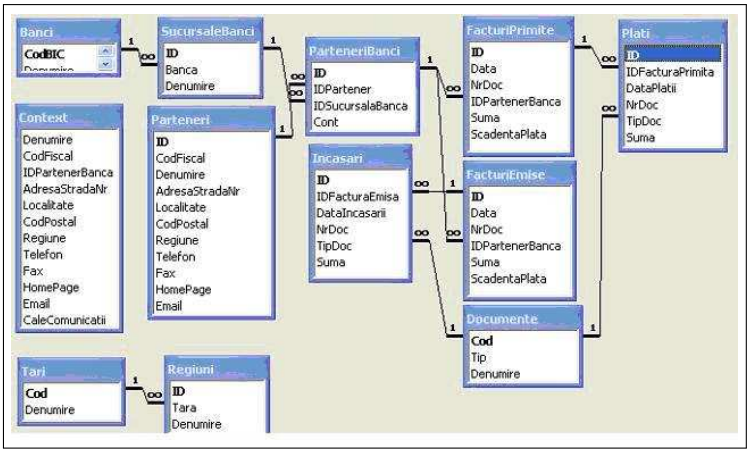


Figure 3: A picture of a window of DB-CRM data base scheme

tions regarding the administration of received and sent e-mails from/to the partners, received and sent faxes and letters from/to the partners. These massages are stored in communication files with the customers or they can be grouped on different criteria and can be displayed on the customer station screen, if requested;

B) Administration of and /or relative public quotas of the enterprise' partners - permits the grouping of partners on type of partners, following the history of their payments to the enterprise. For customers, for example, we can have in view at least the following types:

- faithful customer: buys often from the enterprise, at least once a week;
- good payer customer: pays all the sum until the payment date;
- bad payer customer: pays after the payment date ;
- amnesic customer: forgets to pay;
- new customer: buys for the first time;
- old customer: bought from the enterprise several times.

A large attention given to the data processing regarding the customers is motivated by the fact that the enterprise is interested, first of all, in sales. The classification proposed above, serves as basis for their differentiated treatment, using stimulating or forcing mechanisms. Thus, there can be made discounts for some customer types (maybe even on their birthdays). Credit limits can be introduced (value limit of sales) for other types of customers, as there can be stopped all the sales to some inconvenient types of customers. There can be discussed a possible preferential supply, for some types of customers from the above classification, by a clear specification of the supplying date and the required quantity.

The above quotas, related to the enterprise and to the customers, could be applied to the enterprise suppliers also. For the competition, only public quotas, available on the Internet or on other sources, can be applied.

Regarding the partners information management, related to other criteria than that of accomplishing the payment liabilities toward the firm. We believe that their public quotas could be considered, they are available at Companies Register or published by the professional associations or the specialized organizations on market studies.

3.3 Specific data of Partners subsystem

In order to accomplish this function, in the data base shall be added at least the following information (tables, files or folders):

- partner communications (e-mail, fax etc.) are stored in a folder having the following structure:
  - the root: a way put in the Context table, the WayInfoPartners field (text)
  - every contact partner/person will have a subfolder with two other subfolders:

  - WayInfoPartner - the specific folder to each partner;
    * **Mail**;
      · **In** - received messages;
      · **Out** - sent messages;
      · partners digital photos for contact persons are stored in a folder having the following structure
- the root: a way put in the **Context** table, the **WayContactPerson** field (text);
- every contact partner/person will have an **Img** subfolder that will store the pictures and a Text one that contains unstructured text:
  - **WayContactPerson** - the specific folder of each contact person;
    * **Text** - unstructured text;
    * **Img** - photos.

- other information about the partners, such as :

- partner quotas (faithful, good payer etc.);
- type partner (customer, supplier, competitor);
- date and place of birth etc. (for the partners private persons).
- " information about the competitors (for the competitor partner type) - it is stored like the communications with the partners, that are:
  - **WayInfoPartner** - the specific folder to each partner;
    * **Doc**
      · **Html** - partners' Web pages links or about them;
      · **Text** - unstructured information.

Table 3: Partners Subsystem. New tables in the data base

Table name	**Fields**
PartnersInfo	PartnerID, IDContactPerson, WayInfoPartner, partner type, enterprise quota
Persons	ID, name, date of birth, address, WayContactPerson, phone, email, web page, function, sale quota, supply quota
Quotas	ID, name, organization
PartnersQuotas	PartnerID, QuotaID, quota
Organizations	ID, name, address

The meanings of the included data in the above tables are the following.

PartnersInfo table contains supplementary information for each partner, having the same primary key as the **Partners** table from DB-CRM.

Persons table contains information about the private persons generally speaking, (contact persons). Its key, ID, is associated to the foreign key **IDContactPerson** from **PartnersInfo** table. So, there is **1:1** association between **PartnersInfo** and **Persons** tables.

Quotas table is a list of enterprise' external quotas (independent), and the **PartnersQuotas** table includes the partners' external quotas. Between **Quotas** table and **PartnersQuotas** table there is a relation n:m defined by the foreign key **PartnersQuotas.IDQuota**. Therefore, a quota can be attributed to more partners, and a partner has more quotas, established by different organisms.

Organizations table is the list that includes some information about the non-governmental organizations (in general) that have as duties to make quotas. Between **Organizations** table and **Quotas** table there is a **1:n** relation defined by the foreign key **QuotasOrganization**. A quota is attributed by one organization, and an organization can establish more quotas.

The subsystem functionality includes operations regarding the above described updating tables, the administration of stored information in folders (mail and photos) and synthetic and analytical reports making.

4 Conclusions

In our design we use the iterative and incremental development of the integrated systems of SME, where the first major iteration corresponds to CRM kernel. The further development of the system will be done by using the iterative and incremental development paradigm of integrated information systems as well.

This paradigm has an economic and managerial justification, meaning that any enterprise from a commercial domain must stress the relationship with the partners (customers, suppliers and competitors).

The CRM kernel of the integrated information system has the role to administrate the information about the enterprise' partners and about the commercial relations with them, first of all the following: the issued and received invoices, incomings and payments.

References

[1] V. Lupse, / *Contributions to design and configuration of an information system for small and medium enterprises*, PhD Thesis, "Babes-Bolyai" University, Cluj-Napoca (in Romanian), (2007).

[2] V. Lupse, I. Dzitac, S. Dzitac, E. Valeanu, A Project of Commercial Module of an ERP System for SME, *Informatics in Knowledge Society, Proc. of The Eighth International Conference on Informatics in Economy IE 2007* - May 17-18, 2007, pp. 441-446, (2007).

[3] V. Lupse, I. Dzitac, A design based of object programming of a module of integrated ERP systems of small and medium enterprises, *Acta Univ. Apulensis, Math. and Informatics*, Nr. 10(2005), pp. 275-282, (2005).

[4] V. Lupse, I. Dzitac, A Survey of the ERP Systems for Small and Medium Enterprises, *An. Univ. din Oradea*, Fasc. Electrotehnica, St. Calculatoarelor si sistemelor, pp. 68-75,(2005).

[5] Manolescu M.-J, Re-engineering: a method for re-designing of organization, *AGORA-Studii*, Vol.5 (2005), pp. 185-188, (in Romanian) 2005.

[6] Manolescu A., *Management of public services*, Ed. Universitaria, Craiova, (in Romanian) 2004.

Vasile Lupşe
North University, Baia Mare, Romania
Department of Mathematics and Informatics
E-mail: vasilelupse@yahoo.co.uk

Ioan Dziţac, Adriana Manolescu, Mişu-Jan Manolescu
Agora University, Romania
Economic Informatics Department
E-mail: {idzitac,adrianamanolescu,rectorat}@univagora.ro

Simona Dziţac
University of Oradea, Romania
E-mail: sdzitac@rdslink.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 381-386

The Balance Problem for a Deterministic Model

Mariana Luţă, Luis-Raul Boroacă

Abstract: The study presents the balance problem in the case of a deterministic model, and the necessity of a systematic approach of the enterprise, for a system that is based on the use of economic mechanisms, through artificial intelligence methods. The study presents the balance problem as a problem of minimum distance; it also presents the algorithm and the procedure used to generate the admissible groups associated to a graph.
The balance problem of a production line that assembles a single model of a product and that has deterministic execution periods and a fixed operation rate (of the production line) is a fundamental problem of the management of a flow production line (for processing and assembling).
Keywords: the system approach of the enterprise; the balance problem of production; NP-complete, NP-hard, algorithm,

1 Introduction

The increase of the complexity of processes and phenomena determined the intensification of the research concerning the improvement of the theoretical and practical methods and models used to manage those processes and phenomena at the microeconomic level and to deal with them at the macroeconomic level.

In this context, the concepts of *system* and *systematic* thought represent significant results of present scientific research.

The cell is the basic element of the organism, and the atom is the basic element of matter. In the same way, the enterprise is the fundamental component of the economy.

This "economic atom" (the enterprise) is truly indivisible. The three fundamental economic functions of the enterprise (production, selling and management) can be found at all economic levels: world economy, national economy, economic sectors and enterprise.

2 The system approach of the enterprise

The system approach of the enterprise and the use of an integrated system of resource management offer managers the possibility to have a clear image of their enterprise at any time, and allow them to adopt the right decisions. This way, the enterprise can adapt to the external economic environment, preserving its financial stability.

Modern enterprise offers a significant variety of services; it is adaptable to internal and external factors and the microeconomic decision depends on the alternative possibilities of the market and the uncertain aspects of demand that are more and more difficult to predict. Such an enterprise has a modern and dynamic management that leads to better results in terms of productivity, expenditure, profit, natural environment protection, and social desiderata; all these gives the enterprise competitive advantages on the domestic and international market.

Some of the main goals of a "smart" management system are to reduce the risk, to stimulate creativity and to make people more responsible in the process of decision. We can consider decision to be a process of risk decrease, because it is based on information, experience and ideas that come from many different sources and that can be used in an efficient anticipative way.

The development of cybernetics and of the theory of systems led to a new approach of the enterprise: the enterprise as an industrial cybernetic system. The goal of this industrial cybernetic system is to turn input into output in order to satisfy some social needs. The transformation of resources into products has to be sufficient in terms of quantity and quality; so, that transformation has to be done on the principles of maximum productivity and of minimum consumption of resources.

Two of the performance indicators of an industrial enterprise are *productivity* and *the operation safety of technological equipment* (because they have to protect the worker and the quality of the environment). The modern employee wants a secure job and he or she wants to have an active role in production improvement, to communicate with his or her colleagues and leaders, and to achieve personal career development. Actually, these goals that are more and more important in the last decades can find an explanation in Maslow's pyramid of needs (1954).

A possible solution may be the use of artificial intelligence (based on knowledge) in the management of industrial systems; artificial intelligence methods resemble the processes of the left cerebral hemisphere of the human brain. After 1990, they use artificial neural networks; these neural networks function in a similar way with right cerebral hemisphere of the human brain, and they are used to solve problems that can not be understood at the present level of knowledge. Even so, unpredictable events can occur in practice, due to a certain conjuncture.

Back to 1990, Martin and other authors showed that artificial intelligence and expert systems can be used only for support functions, although they (artificial intelligence and expert systems) had succeeded to solve problems that could not be solved with classic numerical methods; so, renouncing to human evaluation of computerised solutions and results is a big and dangerous error. In 1990, Martin recommends appropriate automation that should integrate technical, human, organisational, economic, and cultural factors. In 1995, Brandt recommends the "dual design" for the systems that are in course of modernisation through technology and automation; according to his approach, they should consider both the technological and the human factors. In 1995, Camarinha-Matos and Afsarmanesh propose balanced automation.

The manufacturing system will be a part of the production system. The manufacturing system will be represented only through certain tasks within the production system, while the production system is the general framework of all activities (both functional and productive activities). A manufacturing task is a particular case of a production task.

The main problem of continuous flow production is the balance of production. The aim of that balance is to cut down production cost, for a certain period of time or for the lifetime of the production line, to distribute tasks correctly, taking in consideration the limits of the system.

The objectives of balance models are: either the efficient use of labor, or the efficient use of equipment, or the efficient use of both (labor and equipment). It is not easy to divide the operations made by machines; because of that, it is more difficult to balance processing operations than assembling operations. In order to avoid imbalance and to assure the operation safety of production lines, they can use work-in-progress production.

In order to balance a production line, they have to observe the following stages:

1. to establish the stages of the technological process;
2. to show the precedence restraints;
3. to point out the possible incompatibility between different stages;
4. to chose a target function;
5. to establish the variation interval of the production-line rate;
6. to group the stages in order to observe the rate of the production line and to optimize the target function.

Generally, they accept that the execution periods of production stages should be considered deterministic. In the case of the production lines where the majority of activities are executed by humans, the execution periods are random variables with known repartition functions.

3 The enunciation of the problem and its complexity

Here, we formulate the balance problem of a production line that assembles one model of a product; the execution periods of production stages are deterministic, and the operation rate of the production line is determined.

We consider that $F = \{1,2,\ldots,n\}$ is the group of production stages; the function $t : F \to \Re_+$ associates the execution time $t(i)$ of the stage i to that stage, according to the model. The execution of the production stages depends on some precedence conditions, and we associate the digraph $G = (F,U)$ to those conditions; we consider that if $x,y \in F$ then $(x,y) \in U$ if and only if the beginning of the stage y depends on the finalisation of the stage x.

Technologically speaking, the digraph will be acyclic if the beginning of each phase does not depend on the finalisation of the stage execution. We consider that $R \in \Re_+$ represents the rate of the production line; we define the rate of the production line as the necessary time for a lot of goods to leave the production line. Considering the technological meaning of rate, we have the following relation: $R \geq t(i), i = 1,2,\ldots,n$. We consider that $Q = \{Q_1,Q_2,\ldots,Q_m\}$ is a group of production stages, and $Q_i \subseteq F, 1 \leq i \subseteq m$; this group of stages observes the following conditions:

$$Q_1 \cup Q_2 \cup ... \cup Q_m = F \tag{1}$$

$$Q_i \cap Q_j = \emptyset, \quad 1 \leq 1 < j \leq m \tag{2}$$

$$T(Q)_j = \sum_{x \in Q_j} t(x) \leq R, \quad 1 \leq j \leq m \tag{3}$$

$$(x,y) \in U, x \in Q_r \text{ and } y \in Q_s \text{ implies that } r \leq s \text{ for each } (x,y) \in U \tag{4}$$

The groups $Qj, 1 \leq j \leq m$ are called working stations. According to the conditions (1) and (2), each stage in F is assigned just to a working station. According to (3), the execution period of each working station should not exceed the rate (the execution period of a working station is the sum of the execution periods of the component stages).

The condition (4) shows the precedence restraints: if the stage production x is executed in the working station Q_r and it precedes the stage y that is executed in the station Q_s, then Q_r will be placed before $Q_s (r < s)$; the stages x and y will be executed in the same station if $r = s$. The shift of the lot of goods from the station Q_j to the station $Q_{j+1}, 1 \leq j \leq n-1$, needs the time period R; the operations in Q_j are over in the period $T(Q_j)$, and then the lot of goods wait for the shift a period that is $R - T(Q_j)$. That waiting period is the dead time of the working station Q_j.

Observing the conditions (1) - (4), the overall dead time of $Q = \{Q_1, Q_2, \ldots, Q_m\}$ is:

$$T^-(Q) = \sum_{j=1}^{m} [R - T(Q_j)] = mR - \sum_{i=1}^{n} t(i) \tag{5}$$

We consider that d is the group of stages $Q = \{Q_1, Q_2, \ldots, Q_m\}$ that observes the conditions (1) - (4). We also consider $Q^* \in \delta$, so that:

$$T^-(Q^*) = \min\{T^-(Q) | Q \in \delta\} \tag{6}$$

It is obvious that Q* observes (6) if and only if:

$$|Q^*| = \min\{|Q| / Q \in \delta\} \tag{7}$$

so, the overall dead time is at the minimum level if the stages are assigned to a minimum number of working stations. The problem of balance in the case of a single model with deterministic execution periods and a fixed rate resides in determining a partition of $Q^* \in \delta$ that observes (7). While partition $Q^* \in \delta$ is a **solution of the problem**,$Q^* \in \delta$ defined through (7) will be an **optimal solution**. If G is acyclic and $R \geq t(i), i \in F$, then $\delta \neq \emptyset$.

We shall suppose that $t(i) \in Z, 1 \leq i \leq n$ and $R \in Z$.

The condition that the execution periods and the rate should be integers is observed, and the value of the overall dead time (as defined through (5)) multiplies with the same amount, without affecting the optimality of a solution. The condition may seem restrictive, but it is not: if some of the numbers would be rational but not integer, they will become integers when R and $t(i), 1 \leq i \leq n$, are multiplied with the lowest common multiple of the denominators.

On the other hand, if either the rate or the execution periods were irrational, the computer would treat them as rational.

We consider EO a given optimization problem and ED an optimization problem associated to EO; so, if $m_0 \in Z_+$, we should decide whether there is $Q = \{Q_1, Q_2, \ldots, Q_m\} \in \delta, m \leq m_0$.

A problem P will be NP-complete if it is NP-hard and if P can be solved in polynomial time through a nondeterministic algorithm.

Definition 1. A problem P will be NP-complete if it is NP-hard and if P can be solved in polynomial time through a nondeterministic algorithm.

We shall demonstrate the NP-completeness of the ED problem. To show that ED is NP-hard, we shall use the fact that the problem BP ("bin packing") is NP-hard: "if there are the numbers $a_1, a_2, \ldots, a_h, B, K \in Z_+$, we shall have to decide whether there is a partition $\{H_1, H_2, \ldots, H_k\}$ of the group $H = \{1, 2, \ldots, h\}$, so that $\sum_{i \in H_j} a_i \leq B, 1 \leq j \leq k$ *and* $k \in K$."

The BP problem is reducible to an ED problem (the BP problem will result from the ED problem if there are no precedence restraints in the ED problem, meaning $U = \emptyset$). Hence, ED is NP-hard.

A nondeterministic algorithm is an algorithm that can have operations with a multiple-choice result: the result of such an operation is not defined in a single way, but there is a specified group of possibilities from which the machinery can choose.

A nondeterministic algorithm is specified when taking into consideration three functions:

1. to choose (A): it means they have to choose an element of the group A;
2. failure: it means the algorithm failed;
3. success: it means successful finish.

A nondeterministic algorithm will fail when there is no group of choices that can lead to success. If every phase of the nondeterministic algorithm has a fixed period of time, then the calculation period necessary to such an algorithm can be defined as the minimal number of phases necessary to finish the algorithm successfully (considering that there is a group of choices that can lead to a successful finish). It is possible to prove that the ED problem can be solved in polynomial time, with a nondeterministic algorithm.

So, the ED problem is NP-complete. Because the ED problem is reducible to an EO problem, it results that EO is NP-hard, too.

In spite of all efforts, they have not succeeded until now to find out a deterministic algorithm for an NP-hard problem; hence, it seems those problems can not be solved in polynomial time.

This aspect can explain the fact that all exact methods proposed to solve the EO problem need exponential time resources, because there is no polynomial algorithm for such a problem nowadays.

4 Algorithms and exact methods of balance in the case of deterministic periods and of a single model. Balance as a problem of minimum distance

If we consider a balance problem with deterministic periods and fixed rate, we shall draw a graph so that solving the problem is reducible to determining the minimum distance between two points of the graph.

In this transformation, the number of graph points increases very fast, according to the number of stages. In spite of the fact that determining the minimum distance between two graph points can be done in polynomial time (according to the number h of graph points), obtaining an optimal solution of the balance problem needs a lot of time and space resources in this case.

We shall consider: the group of production stages, $F = \{1,2,\ldots,n\}$; the acyclic digraph of the precedence restraints, $G = (F,U)$; the execution period of the stage i, $t(i) \leq R, i = 1,2,\ldots,n$(R is the rate of the production line).

If $x \in F$, then be $\pi(x)$ the group of the direct predecessors of the point x in G, and $\sigma(x)$ the group of the direct successors of the point x in G; so, $\pi(x) = \{y/y \in F,(y,x) \in U\}$ and $\sigma(x) = \{z/z \in F,(x,z) \in U\}$. If $E \in F$, then be $\pi(E) = \bigcup_{x \in E} \pi(x)$ the group of the direct predecessors of the group E.

Definition 2. A point $x \in F$ is a direct successor of the group $E \in F$ (where $x \notin E$) if $\pi(x) \neq \varphi$ and p(x) $\pi(x) \in E$.

A phase that does not belong to E will succeed E directly if it admits direct preceding phases that do not belong all to E. We shall consider that $\lambda(E)$ is the group of phases that succeed $E \in F$ directly.

Definition 3. . A point group E of the graph G is admissible if $\pi(E) \in E$.

A characteristic of an admissible point group is that if it includes a certain phase, it will include all phases that precede that phase directly or indirectly.

We shall consider the family $F = \{E/E \in F, Eadmissible\}$ of the admissible groups of the digraph $G = (F,U)$. It is obvious that $F \in F$ and, conventionally, $\varphi \in F$.

The following procedure allows generating all the admissible groups associated to a digraph.

Procedure GENADMI(G,Fa)

// G=(F,U) - acyclic digraph ;

$F_k = \left\{E_1^k \ldots E_{s_k}^k\right\}$ - the family of the admissible groups generated in the phase k. //

1. begin $F0 = \varphi$; // phase 1 //
2. $F1 = \{E/E \in F, \pi(E) = \varphi\}$;
3. $F_s = F_0 \cup F_1$;
4. for $k = 2$ step 1 until $Fk - 1 = \{F\}$ do // phase k - generating the group Fk //
5. begin $Fk = \varphi$;
6. for $h = 1$ step 1 until sk-1 do
7. begin $B = \lambda(E_h^{k-1}) \cup \{x/x \in E_h^{k-1}, \pi(x) = \varphi\}$;
8. $F_k = F_k \cup \{E_h^{k-1} \cup Y/Y \subset B\}$;
9. end;
10. Fa=F0 U Fk;

Theorem 4. *If the digraph G is acyclic then:*

(i) *the algoritm ends ;*

(ii) *Fa = F*

We consider $F = \{X_0, \ldots, X_r\}$ where $X_0 = \varphi$ si $X_r = F$. We shall make the digraph $\Gamma = (R, U)$, where $R = \{0, 1, \ldots, r\}$ is the point group, and, if i, $j \in R$, then $(i, j) \in U$ if and only if $Xi \subset Xj$ and $T(X_j \backslash X_i) = \sum_{x \in X_j \backslash X_i} t(x) \leq R$.

We consider that $1 : U \to \Re^+$ is the function that associates the distance $l(i, j) = R - T(X_j \backslash X_i)$ to each arch $(i, j) \in U$.

We consider that $d = [i_0, i_1, \ldots, i_k, i_{k+1}]$ (where $i_0 = 0, i_{k+1} = r$) is a line from the point 0 to the point r in the digraph Γ. Considering that $t(i) \leq R, i \in F$ and that G is acyclic, it results that there is such a line (according to [Teorem 4]).

The length of the line d is $L(d) = \sum_{j=0}^{k} l(i_k, i_{j+1})$

If $X_{i_0}, X_{i_k}, X_{i_2}, ..., X_{i_k}, X_{i_{k+1}+1}$ are the admissible groups associated to the points $i_0, i_1, i_2, \ldots, i_k, i_{k+1}$ of the line d, we shall consider $Q(d) = \{Q1, \ldots, Q_{k+1}\}$, where $Q_j = X_j \backslash X_i$ for $j = 1, 2, \ldots, k+1$.

The following theorem shows the link between the group d of the solutions of the balance problem (in the case of deterministic periods and a fixed rate) and, respectively, the group D of the lines from the point 0 to the point r in the digraph Γ.

Theorem 5. *If G = (F,U) is acyclic and $t(i) \leq R$ for any $i \in F$ then*

$$\delta = \{Q(d)/d \in D\}$$

We can notice that the length of a line $d^* \in D$ is minimal if and only if it includes a minimum number of arches. So, finding the optimal solution of the balance problem in the case of deterministic periods and a fixed rate resides in finding the line from the point 0 to r that should include the minimum number of arches.

Considering that, it does not result it is necessary to generate all admissible groups and to draw the graph completely; the number of the graph points increase very fast together with the number of phases. A line from 0 to r that has a minimum number of arches can be determined this way:

Firstly, we shall determine the group X_1 of all points h in Γ that observe the relation $T(X_h) \leq R\, T(X_h) \leq R, X_h \in F$, and the group $Ul = \{(0, h)/h \in X_1\} \subset U$.

When $X_k (k \geq 1)$ determined, we should determine (in the phase k+1) the group X_{k+1} that includes (for every $h \in X_k$) all the points j that observe the relation $X_h \subset X_j, T(X_j, X_h) \leq R, X_j \in F$ and $(h, j) \in U_{k+1}$. The procedure should go on until we reach the point r in G. Of course, the most unfortunate case is when we reach the point r just in the phase n.

5 Conclusions

In order to achieve the efficiency goals, a production activity that involves significant material and human resources (in order to make an important amount of goods) must be organised scientifically.

To balance production lines is a way to increase economic efficiency.

The balance problem is a kind of problem for which they were not able to imagine a suitable algorithm, because of the limits of time resources. This is the reason for which it seems they should develop heuristic methods.

References

[1] C. Barbulescu, *Managementul productiei industriale*, Editura Sylvi, Bucuresti, 1997.

[2] E. S. Buffa, *Conducerea moderna a productiei*, vol. I si II, Editura Tehnica, Bucuresti, 1993.

[3] F. Luban, V. Dumitru, *Contributii la modelarea si rezolvarea unor probleme de repartizare a productiei pe masini prin programare matematica, Cibernetica*, Editura Academiei , Bucuresti, 1982.

[4] J. L. C. Macaskill, *Production line balances for mixed-model lines*, Management Science, 1972.

[5] I. Popescu, D. Radulescu, *Modelarea sistemelor de productie*, Editura Tehnica, Bucuresti, 1986.

[6] F. M. Tonge, *Assembly line balancing using probabilistic combinations of heuristic*, Management Science, 1965.

[7] H. J. Zimmermann, M. G. Sovereign, *Quantitative models for production management*, New Jersey, 1974.

[8] L. Wester, M. D. Kilbridge, *Heuristic line balancing - a case*, Journal of Industrial Engineering, 13, 3, 1962.

Mariana Luţă, Luis-Raul Boroacă
"Emanuil Gojdu" Economic College of Hunedoara, Romania
1bis, "Alexandru Vlahuta" Street
E-mail: inf_ml@yahoo.com;luisboroaca@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 387-392

Equilibrate Cutting Trees

Ioan Maxim, Ioan Tiberiu Socaciu-Lendvai

Abstract: The total or partial covering of plane surfaces with a pre-established set of forms, can be applied in many domains. The problem of covering the surface is followed by the one of creating an order and cutting the forms that cover the surface. The paper presents an algorithm to order the cutting process by using equilibrate cutting trees and considering the conditions and the restrictions imposed by the wood industry.
Keywords: cover of plane surfaces, cutting diagrams, cutting trees, equilibration components

1 Introduction

In the wood industry, the technological cutting restrictions impose a specific cutting succession. In [3] I presented an hierarchical classification algorithm of the cuts, which can satisfy the technological cutting restrictions too.

As in the short-serried production the cutting way is established manually, an important objective is the growth of the work production through a better manipulation of the cutting surfaces. The algorithm presented in [1] produces a succession of cuts, which is not unique. The succession of the cuts can be unique only by imposing some restrictions, which can lead to the equilibration of the cut resulting components, in the cut conditions.

One of the important issues in the cutting diagram creation is the diagram-cutting problem. A cutting diagram, created by a special algorithm, might not be able to be cut, meaning that there is no cuts succession which could make possible the surface cutting in the given forms and technological restrictions. The restrictions imply the existence of an edge to cut, in any cutting moment.

The algorithm presented in [2] decides if the cutting diagram is accepted or not, and if the diagram is accepted, a succession of cuts will be generated. The algorithm takes the solution of constructing a not-oriented graph that should have topological sorted knots. If the topological sorting operation ends up successfully, the cutting diagram is accepted.

2 The cutting binary trees creation

The algorithms that generate cut able diagrams are presented in [1]. One aspect in question is that of establishing the cuts succession in some restricting conditions, which can influence positively some productivity indicators.

The Optimum lexicographic[1] algorithm that creates cutting diagrams suggests a cuts succession, which can be improved for equilibrating the cutting resulted surfaces[4]. The cutting diagram generation can insert a binary cutting tree [3]. For the cutting diagram in figure no. 1a), the binary cutting tree is generated by the next algorithm:

- a cut inserts an intermediary node in the binary tree;
- the ending nodes in the graph are waste forms and surfaces; - every cut will be associated to a direction (horizontally or vertically);
- a cut always generates two surfaces, which in the cutting diagram appear on the left and on the right, or up and right.

In the tree construction the cutting-node is associated with the left under-tree corresponding with the left or up surface, and the right under-tree with the right or down surface.

In order to simplify the cutting tree, we agree that a final node, which is a form, be at the same time a group of forms of the same type, obtained from a diagram generation procedure. The forms and the groups are marked with capital letters and the cuts with numbers. (fig. 1)

The binary tree in figure no. 2 is generated corresponding to the cutting diagram in figure no. 1.

The crossing of the tree reflection leads to the next cuts succession: 1, 3, 6, 9, 11, 14, 16, 17, 20, 21, 22, 2, A, 4, 5, B1, B, 7, 8, C1, C, 10, D, 12, 13, E1, E, 15, F1, F, G, 18, 19, H1, H, I, I1, 23, I11.

The mirroring operation can be avoided by realizing a tree crossing by the root-write-left crossing succession. This method will be named the d-preorder crossing in the text below.

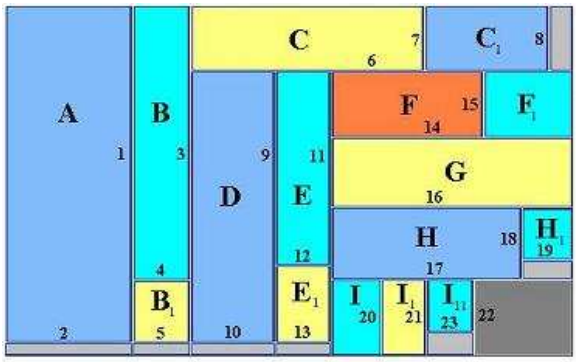


Figure 1: Cutting diagram

This cuts succession correspond entirely to the cutting prescribing, that forecast first the large surfaces cutting, and then the cutting of the forms in those surfaces. Such a procedure contributes to the growth of work productivity. The work routine is helping by periodically relaxing the worker.

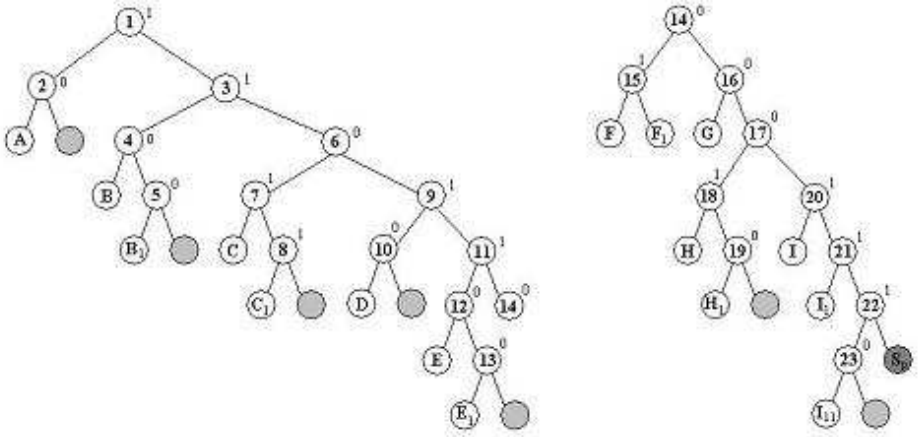


Figure 2: The binary cutting tree

Another condition for the productivity growth, if the operation implicates manual acts, is the one linked to the surface equilibration. Therefore, the acceptable cut for equilibrium must be found, if possible.

3 The binary cutting trees equilibration

Normally, in order to find the cutting equilibrium, a search operation is needed. The cutting tree resulted after the last step is strongly unbalanced on the right. A tree equilibration followed by a preorder crossing leads to a cutting succession that balances the cut surfaces.

Each cut can be associated to a direction; 0 - horizontal or 1-vertical. In the table below (figure 3.), which was made for the cutting diagram in figure no. 1, the first line is counting the cuts, and the second one the directions.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	0	0	1	0	1	1	1	0

Figure 3: Cutting directions

Definition:*A cutting tree is considered to be unbalanced if it has an under-tree in which the succession of right descendents of the under-tree's root, marked with the same direction sign, are followed by a right descendent with another direction mark.*

The balancing of the cutting tree will be realized by a right side rotation of a pivot node [5].

The balancing, respecting the definition conditions and the notation in figure no. 4a), consists in some left side rotations, by the next step:

- the pivot node y is established to be the last right descendent of the under tree root from the same direction mark sequence;
- this y node becomes the under tree root;
- his x father becomes his left son;
- his left son becomes the left son of his father.

The transformation described in the algorithm is represented in figure no. 4b).

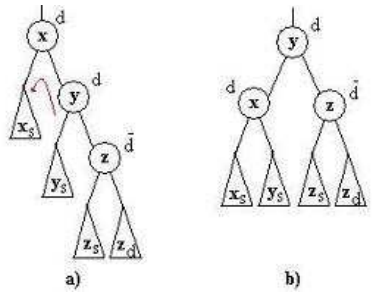


Figure 4: Left rotation

For the cutting tree in figure no. 2, the algorithm application identifies the nodes no. 3, 11 and 17 as balance pivot nodes. The result is the equilibrated cutting tree in figure no. 5.

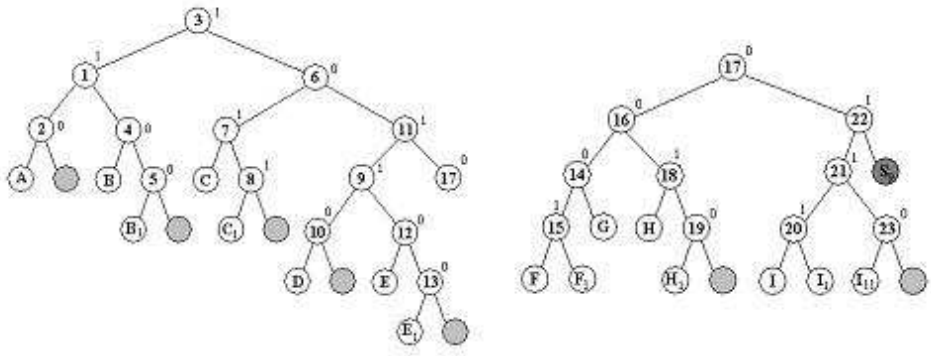


Figure 5: Equilibrated cutting tree

A crossing in d-preorder of the equilibrated cutting tree in figure no. 5 leads to the next nodes succession: 3, 6, 11, 17, 22, 21, 23, I11, 20, I, I1, 16, 18, 19, H1, H, 14, G, 15, F1, F, 9, 12, 13, E1, E, 10, D, 7, 8, C1, C, 1, 4, 5, B1, B, 2, A. The next nodes succession is obtained by eliminating the ending nodes: 3, 6, 11, 17, 22, 21, 23, 20, 16, 18, 19, 14, 15, 9, 12, 13, 10, 7, 8, 1, 4, 5, 2.

This cut order corresponds to a cutting principle according to which first surfaces to be cut at first are the big ones, then the same logic is applied to the remaining surface. This cutting principle is practically sustained by the store space economy, near the cutting equipment.

The same type forms group must be treated differently; they were noted down with capital letters in the given example.

A group can contain an even or an uneven number of forms, which generates an uneven or even number of cuts, parallel at the same distance one from another. The cutting direction is given by the node father in the cutting tree. The cuts order counted 1, 2, ..., n is not given by the cut(u,v), where u+1 and v-1 represent the order number of the first and last cuts for the given surface:

procedure cut(u,v)
if (v-u)>1 then m=[(u+v)/2] write m
cut(u, m)
cut(m, v)

This modality can be applied for any surface that contains same direction cuts. It is best to use the binary principle for cuts ordering (the first cut is the one most near to the cutting surface center), for the best surface equilibrium.

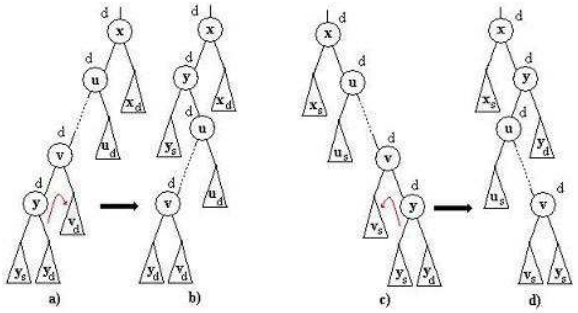


Figure 6: Right side rotations and left side rotations

The best result is a completely equilibrated cutting tree, even form this point of view, and it can be obtained by a new equilibrating operation on the tree that was formerly equilibrated.

The cutting tree balance is completed by some right side rotations, by those steps:

- the pivot node y is established as being the middle node for a 3 left descendents sequence with the same direction mark, in an x-root under tree;
- we note with u-x's left son, and with v-, y node's father;
- the y node becomes x's left son;
- the u node becomes y's left son;
- the y's right son becomes v's left son or of some left rotations (fig.6) which assume the next steps:
- the y node is established to be the middle node from a sequence of three right-descendents, having the same direction mark in an x-root under tree;
- we note with u-x's right son, and with v- the father of the y node;
- the y node becomes x's right son;
- the u node becomes y's right son;
- the Y's left son becomes v's right son.

The right side rotation, described in the algorithm about the x-root under tree, in figure no. 6a) is drawn in the figure no. 6b).

The cutting tree in figure no. 8a) results from the cutting diagram in figure no. 7. In this tree, the nodes no. 4 and 9 are identified to be necessary for equilibrating. The result is the tree in figure no. 8b). The tree obtained in this mode has three unbalanced under-trees-root no. 1, 6, and 11, which will be rotated on the right side. In the end, a right side rotation will be applied to the root 7 under -tree. The result is a perfectly equilibrated tree (fig. 8c).

The d-preorder crossing of the perfectly equilibrated cutting tree leads to the next node succession: 4, 9, 10, 11, L, K, J, 7, 8, I, H, 5, 6, G, F, E, 2, 3, D, C, 1, B, A. If the leaf nodes labels are eliminated from the row, the succession becomes: 4, 9, 10, 11, 7, 8, 5, 6, 2, 3, 1.

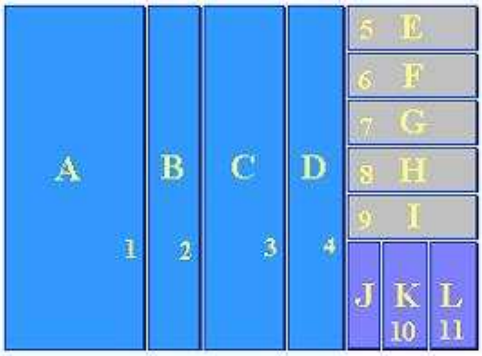


Figure 7: Cutting diagram

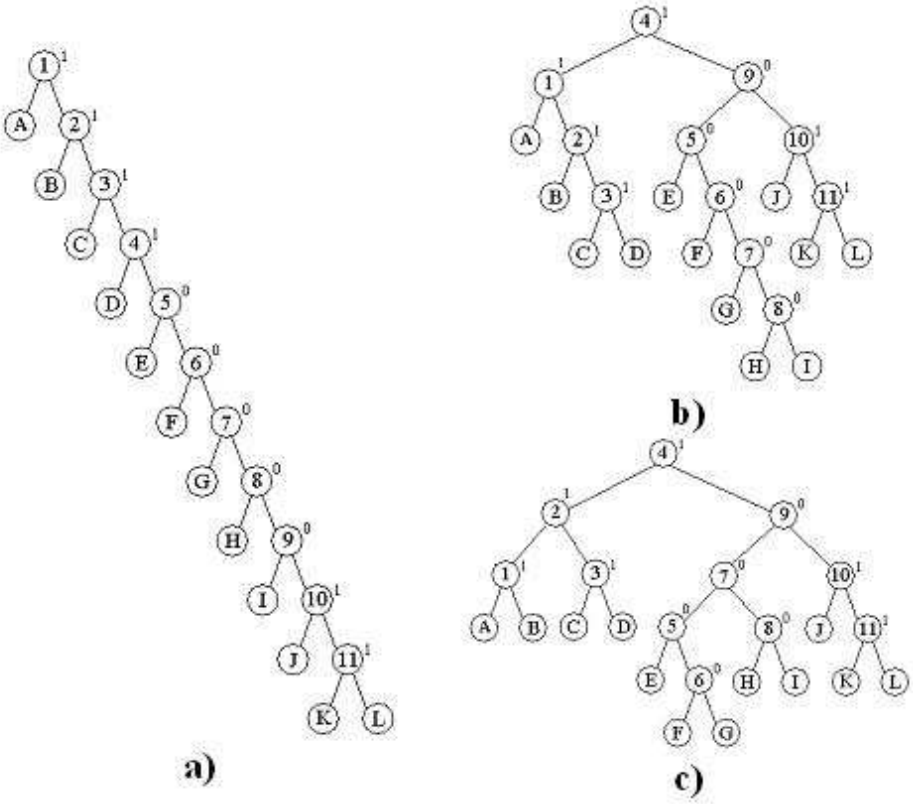


Figure 8: Cutting tree

4 Conclusions

The big number of domains where this cutting theory can be applied makes it possible to be discussed in many manners. That is why, taking into consideration the furniture industry restrictions, the issues that were mainly discussed in were: the cutting diagram elaboration, resolving the cutting diagram problem, cuts hierarchy, the surface equilibration by modifying the cuts order. The cutting tree notion was introduced. It generates the cuts precedence graph that can solve the cutting problem by one surface crossing. Another important author's contribution to the surface equilibration problem solving is the balancing algorithm creation. The algorithm operates on the cutting tree for obtaining a new cutting hierarchy that will lead, as much as possible, to equilibrate surfaces [4]. The economic importance of this algorithm is shown by the human operator productivity for manual cutting of the guide marks.

References

[1] I. C. Maxim, "Metoda acoperirilor succesive pentru optimizarea algoritmului de acoperire a suprafetelor planare fara restrictii," *Tehnologii Informationale, Seminarul "Sisteme distribuite", Univ. "Stefan cel Mare"*, Suceava, pp. 138-144, 2003.

[2] I. C. Maxim, "Cutting diagrams cut up for plane surface with restrictions," *Procesare distribuita, Seminarul "Sisteme distribuite", Univ. "Stefan cel Mare"*, Suceava, pp. 99-102, 2005.

[3] I. C. Maxim, "Clasificarea ierarhica a taieturilor," *Analele stintifice ale Universitatii de Stat din Moldova, seria "Stiinte fizico-matematice"*, Chisinau, pp. 132-134, 2005.

[4] I. C. Maxim, "The classification of the cuts through the equilibration of the components," *SINTES, The International Symposium on Systems Theory, Software Engineering; XII-th edition*, Vol. III, Romania, Craiova, pp. 620-623, 2005.

[5] I. C. Maxim, E. Mateescu , *Arbori,* Editura "Tara fagilor", Suceava, 1996.

Ioan Maxim
"Stefan cel Mare" University of Suceava
Teacher Training Department
Universitatii, 13, 720229, Suceava, Romania
E-mail: maximioan@yahoo.com

Ioan Tiberiu Socaciu-Lendavai
"Stefan cel Mare" University of Suceava
Economic Sciences and Public Administration
Universitatii, 13, 720229, Suceava, Romania
E-mail:tibisocaciu@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 393-395

The Deutsch-Josza's Algorithm for n-qudits

Gabriela Mogoş

Abstract: Deutsch-Josza's algorithm, as all known quantum algorithms that provide exponential speedup over classical systems do, answers a question about a global property of a solution space. This paper describes the generalization of the Deutsch-Josza algorithm to n d-dimensional quantum systems or qudits.
Keywords: quantum dits, quantum algorithms.

1 Introduction

A qudit is a general state in a d-dimensional Hilbert space H_d i. e. $|\Psi\rangle = \sum_{m=0}^{d-1} c_m |m\rangle$, which reduces to $|\Psi\rangle = c_0|0\rangle + c_1|1\rangle$, for the qubit case space.

An n-qudit is a state in the tensor product Hilbert space. The computational basis of H is the orthonormal basis given by the d^n classical n-qudits:

$$|m_1\rangle \otimes |m_2\rangle \otimes ... \otimes |m_n\rangle = |m_1 m_2 ... m_n\rangle \tag{1}$$

where $0 \leq m_n \leq d-1$.

The general state in H is a superposition:

$$|\Psi\rangle = \sum \Psi_{m_1 m_2 .. m_n} |m_1 m_2 ... m_n\rangle$$

where $||\Psi||^2 = \sum |\Psi_{m_1 m_2 .. m_n}|^2 = 1$.

We say Ψ is *decomposable* when it can be written as a tensor product of qudits:

$$|m_1 m_2 ... m_n\rangle = |m_1\rangle \otimes |m_2\rangle \otimes ... \otimes |m_n\rangle = \bigotimes_{i=0}^{n} |m_i\rangle = |m\rangle$$

where $|m_i\rangle$ is a general state in a d-dimesional Hilbert space for one qudit.

2 The Deutsch-Josza's algorithm for n-qudits

The algorithm Deutsch-Josza are generalize Deutsch's algorithm and the function is: $f : \{0,1,...,d-1\}^n \longrightarrow \{0,1,...,d-1\}$.

Taking matters the fact that exist n-qudits as input data, we put a global problems if the function $f(x)$ is constant or balanced. If the function will be balanced, then mean that the entire exit will be 0 for exactly half of the inputs.

Classic speaking that represent the evaluation of the function $f(x)$ for much more that half of inputs, because we must see with certitude when the function is balanced and when is constant.

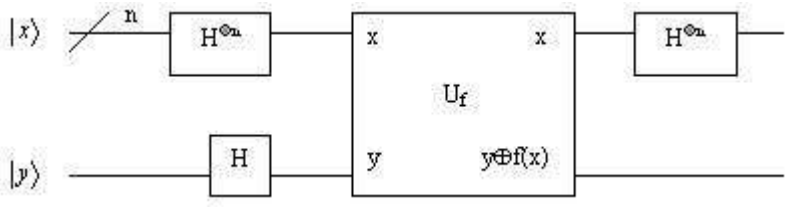


Figure 1: Deutsch-Josza's algorithm

Let us analyze this circuit now.

1. First we need to apply Hadamard gates H, to register with n qudits. We can now apply the Hadamard operator to each qudit of the product state:

$$H^{\otimes n} |m\rangle = H^{\otimes n}(|m_1\rangle \otimes |m_2\rangle \otimes ... \otimes |m_n\rangle) =$$

$$= H^{\otimes n}|m_1\rangle \otimes H^{\otimes n}|m_2\rangle \otimes ... \otimes H^{\otimes n}|m_n\rangle \tag{2}$$

In general, the Hadamard operator acting on a single qudit of dimension d is defined as:

$$H|x\rangle = \frac{1}{\sqrt{d}}\sum_{y=0}^{d-1}(-1)^{x\cdot y}|y\rangle \tag{3}$$

However, we generalize the latter to $H^{\otimes n}$ the notation pays off since the above form can immediately be generalized by summing over all possible combinations of qudit basis states, i.e., over all n-qudit states

$$H^{\otimes n}|x\rangle = \frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{y_i=0}^{d-1}(-1)^{x_i\cdot y_i}|y_i\rangle$$

where

$$x_i \cdot y_i = x_1y_1 + x_2y_2 + ... + x_ny_n$$

For register with n qudits, the equation (3) become:

$$(\frac{1}{\sqrt{d^n}}\sum_{y_1=0}^{n}\sum^{d-1}(-1)^{x_1\cdot y_1}|y_1\rangle)\otimes(\frac{1}{\sqrt{d^n}}\sum_{y_2=0}^{n}\sum^{d-1}(-1)^{x_2\cdot y_2}|y_2\rangle)\otimes ...\otimes(\frac{1}{\sqrt{d^n}}\sum_{y_n=0}^{n}\sum^{d-1}(-1)^{x_ny_n}|y_n\rangle) =$$

$$= \frac{1}{\sqrt{d^{n^2}}}\sum^{n}[(\sum_{y_1=0}^{d-1}(-1)^{x_1\cdot y_1}|y_1\rangle)\otimes(\sum_{y_2=0}^{d-1}(-1)^{x_2\cdot y_2}|y_2\rangle)\otimes ...\otimes(\sum_{y_n=0}^{d-1}(-1)^{x_ny_n}|y_n\rangle)] =$$

$$= \frac{1}{\sqrt{d^{n^2}}}\sum_{i=0}^{n}(\bigotimes_{i=0}^{n-1}\sum_{y_i=0}^{d-1}(-1)^{x_i\cdot y_i}|y_i\rangle) = \bigotimes_{i=0}^{n-1}[\frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{y_i=0}^{d-1}(-1)^{x_i\cdot y_i}|y_i\rangle] = |\Psi_1\rangle$$

More generally, given $|\Psi\rangle = \bigotimes_{i=0}^{n-1}|\Psi_i\rangle$, where each $|\Psi_i\rangle \in H_d$, the state $H^{\otimes n}|\Psi\rangle$ can be computed in linear time by:

$$H^{\otimes n}|\Psi\rangle = \bigotimes_{i=0}^{n-1}H|\Psi_i\rangle$$

2. In next step of the Deutsch-Josza algorithm, we should evaluate the U_f operator effect. The operation of U_f gate is completely defined by its action on the computational basis for each qudit:

$$|x\rangle|y\rangle \xrightarrow{U_f} |x\rangle|y \oplus f(x)\rangle$$

where $|x\rangle$ and $|y\rangle \in \{|0\rangle, |1\rangle, ..., |d-1\rangle\}$ denote the state of control and auxiliary qudits.
Using U_f operator, we now transform the n-qudits of the upper lines and 1-qudit of the lower line, as:

$$[\bigotimes_{i=0}^{n-1}(\frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{x_i=0}^{d-1}(-1)^{z_i\cdot x_i}|x_i\rangle)](\frac{1}{\sqrt{d}}\sum_{y=0}^{d-1}(-1)^{t\cdot y}|y\rangle) \xrightarrow{U_f} [\bigotimes_{i=0}^{n-1}(\frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{x_i=0}^{d-1}(-1)^{f(x)}(-1)^{z_i\cdot x_i}|x_i\rangle)]$$

$$(\frac{1}{\sqrt{d}}\sum_{y=0}^{d-1}(-1)^{t\cdot y}|y\rangle) = [\bigotimes_{i=0}^{n-1}(\frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{x_i=0}^{d-1}(-1)^{f(x)+x_i\cdot z_i}|x_i\rangle)](\frac{1}{\sqrt{d}}\sum_{y=0}^{d-1}(-1)^{y\cdot t}|y\rangle) = |\Psi_2\rangle$$

3. Finally we apply another $H^{\otimes n}$ transform to obtain:

$$|\Psi_3\rangle = \sum^{n}[\bigotimes_{i=0}^{n-1}(\frac{1}{\sqrt{d^n}}\sum_{i=0}^{n}\sum_{x_i=0}^{d-1}(-1)^{f(x_i)+x_i\cdot z_i}|x_i\rangle)](\frac{1}{\sqrt{d}}\sum_{y=0}^{d-1}(-1)^{y\cdot t}|y\rangle)$$

We measure the probability amplitude of $x_i = |m_i\rangle^{\otimes n}$.

For *constant* $f(x_i)$, the sum over x_i is independent of x_i and $x_i \cdot z_i$ must also be equal to zero and hence $(-1)^{x_i\cdot z_i + f(x_i)}$ is either -1 or $+1$ for all values of x_i, where -1 holds for $f(x_i) = 1$ and 1 holds for $f(x_i) = 0$.

In this case the amplitude for $x_i = |m_i\rangle^{\otimes n}$ is:

$$\overset{n}{\otimes}(\pm\sum_{x_i=0}^{d-1}\frac{1}{\sqrt{d^n}}) = \pm 1$$

since $|\Psi_3\rangle$ is normalized to 1 and the amplitude of $x_i = |m_i\rangle^{\otimes n}$ already gives probability 1, there can be no other component in $|\Psi_3\rangle$, all other amplitudes must be zero. Hence when we measure the first n qudits in the query register, we will obtain a zero $(|0_i\rangle^{\otimes n}$.

If $f(x)$ is *balanced* then $(-1)^{x_i \cdot z_i + f(x_i)}$ will be $+1$ for some values of x_i and -1 for other values of x_i. The amplitude of the all states $x_i = |m_i\rangle^{\otimes n}$ is then:

$$\overset{n}{\otimes}(+\sum_{x_{i_1}=0}^{d-1}\frac{1}{\sqrt{d^n}} - \sum_{x_{i_2}=0}^{d-1}\frac{1}{\sqrt{d^n}}) = 0$$

where x_{i_1} is the set of x_i's such that the function $f(x_i)$ has a plus sign and x_{i_2} is the set of x_i's where $f(x_i)$ has a minus sign.

We say that f has a balanced parity when an even value for exactly half of x_i and a odd value for the other half.

References

[1] Vahid Karimipour, Alireza Bahraminasab,*Quantum key for d-level systems with generalized Bell states*, Physical Review A. Vol. 65, 052331, 2002.

[2] Jamil Daboul, Xiaoguang Wang and Barry C. Sanders, *Quantum gates on hybrid qudits*, arXiv:quant-ph/0211185, 2002.

[3] G.Johansson, *Quantum Algorithms-Lectures in Quantum Informatics*, Quantum Physics, Sweden, 2005.

[4] Faisal Shah Khan, Marek M.Perkowski, *Synthesis of ternary quantum logic circuits by decomposition*, arXiv:quant-ph/05111041, 2005.

Gabriela Mogoş
University Al.I.Cuza Iasi
Computer Science Department
Carol I no.11, 700088, Romania
E-mail: gabi.mogos@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 396-401

Designing appropriate schemes for the control of fed-batch cultivation of recombinant *E.coli*

Saleh Mohseni, Ahmad Reza Vali, Valiollah Babaeipour

Abstract: Fed-batch fermentation processes are common methods for producing biologic recombinant cells from different microorganisms. Model-based control of bioprocesses is a difficult task due to the challenges associated with bioprocess modelling and the lack of on-line measurements. This paper aimes designing some feeding control approaches in fed-batch cultivation of high cell density E.coli producing recombinant proteins. For this purpose first a neural network self-tuning PI controller is designed and some advantages and disadvantages of this controller are denoted. Then an optimal controller is suggested for the process and a feedback linearizing controller is used afterwards in order to make optimal controller robust against disturbances and uncertainties. The introduced controller is capable of maintaining the proper specific growth rate for attaining the maximum amount of biomass and recombinant protein production efficiency.
Keywords: fed-batch cultivation, NN based self-tuning PI controller, optimal control, feedback linearizing controller.

1 Introduction

Nowadays many proteins are produced by genetically modified microorganisms. Since the physical, biochemical and genetical properties of the bacterium Escherichia coli are known better, they are the most generic host microorganisms used for the production of recombinant proteins.

The successful and economical run of recombinant protein production is quite dependant on achieving the maximum performance of the protein production. Fed-batch processes are the most current and appropriate way of increasing the production [2, 3]. The successful run of this process is ensured by appropriate control of the feeding rate. In other words underfeeding causes productivity loss and starvation while overfeeding leads to carbon nutrient accumulation or by-product formation, such as acetate [1, 4]. Consequently most researches reported in this area are devoted to feeding control approach.

Control of bioprocesses is a delicate task due to at least two reasons:

- The process complexity, nonlinearity and non-stationarity which make modeling and parameter estimation particularly difficult;

- The scarcity of on-line measurements of the component concentrations (essential substrates, biomass and products of interest)[8].

So far many control methods have been proposed but due to drawbacks associated with each one, none of them is efficient enough to serve as a general method for fed-batch process.

Therefore according to experimental data obtained from previous researches, here a self-tuning PI controller based on neural network is designed and an optimal controller is surveyed the results of which will be discussed afterwards.

1.1 Process Description

Current process is a bioreactor which operates in fed-batch mode. There are two inputs for the system. They are carbon feed rate (F) for nourishing biomass and oxygen (the air) for respiration[1]. The feeding rate is controlled by tuning the feeding pump speed that is a peristaltic pump, while the air flow input can be controlled by changing the stirrer speed (N) or increasing the saturation concentration of dissolved oxygen tension ,DO^*, which is done by injecting pure oxygen. In this study the oxygen and peripheral parameters such as PH and temperature are kept constant and the only control parameter is the feed flow rate.

The process model was constructed based on mass and energy balances of the bioreactor. Identification of the process was done off-line by utilizing genetic algorithm and PSO with the help of experimental data gathered from previous runs[6, 9].

1.2 Process model

The mass balance equations of a fed-batch bioreactor are[9]:

$$\begin{aligned}
\frac{dX}{dt} &= \mu X - \frac{F}{V}X \\
\frac{dS}{dt} &= -\frac{1}{Y_{S/X}}\mu X + \frac{F}{V}(S_{in} - S) \\
\frac{dA}{dt} &= \frac{1}{Y_{A/X}}\mu X - \frac{F}{V}A \\
\frac{dDO}{dt} &= -\frac{1}{Y_{DO/X}}\mu X + k_L^{DO}a.(Do^* - DO) - \frac{F}{V}DO \\
\frac{dCO_2}{dt} &= -\frac{1}{Y_{CO_2/X}}\mu X + k_L^{CO_2}a.(CO_2^* - CO_2) - \frac{F}{V}CO_2 \\
\frac{dV}{dt} &= F
\end{aligned} \tag{1}$$

In which X is biomass concentration $[g/L]$; S - glucose concentration $[g/L]$; DO - dissolved oxygen tension $[\%]$; CO_2 - carbon dioxide concentration $[\%]$; DO^* - saturation concentration of dissolved oxygen tension $[\%]$; CO_2^* - saturation concentration of carbon dioxide $[\%]$; F - feed rate $[L/h]$; V - bioreactor volume $[L]$; S_{in} - glucose concentration of feed $[g/L]$; μ - specific rate of biomass growth $[g\ g^{-1}L^{-1}]$; $Y_{S/X}$- yield coefficient for biomass $[g\ g^{-1}]$; $Y_{A/X}$ - yield coefficient for acetate $[g\ g^{-1}]$; $Y_{DO/X}$ - yield coefficient for dissolved oxygen tension $[g\ g^{-1}]$; $Y_{CO_2/X}$ - yield coefficient for carbon dioxide $[g\ g^{-1}]$;$k_L^{DO}a$ - volumetric oxygen transfer coefficient $[h^{-1}]$;$k_L^{CO_2}a$ - volumetric carbon dioxide transfer coefficient $[h^{-1}]$. Kinetics of the considered growth is as follows [8]:

$$\mu = \mu_{max}\frac{S}{S+k_S}\cdot\frac{k_i}{A+k_i} \tag{2}$$

This equation is based on Monod model and an inhibition term of acetate (in growth of cells) is added to it.

2 Control strategies

Fed-batch processes have two working stages. The first stage is called batch in which biomass grows by means of the substrate(carbon source) in the medium and no external feed is added to the bioreactor. Then after exhausting the carbon source which causes sudden increment of the oxygen concentration as well as sudden decrement of specific growth rate, the batch stage is finished and the next stage is commenced that is called fed-batch cultivation during which glucose feed is added to the culture and the process is nourished by control strategies.

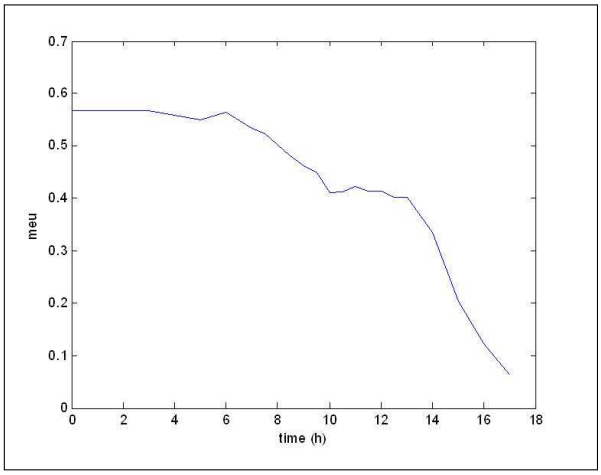


Figure 1: reference specific growth rate μ_{ref}

The scarcity of on-line measurements of the component concentrations makes the use of observers inevitable. So, here a sequential observer is utilized which estimates μ and biomass concentration sequentially from measurement of OUR[7]. The observer's parameters are tuned so that it's stability is guaranteed.

At first a practical setpoint, obtained empirically in the lab, is used from which the maximum concentration of biomass reported ever have been gained[2, 3]. Figure(2) shows the practical reference of μ used as setpoint.

2.1 Neural network based self-tuning PI controller

Neural network based self-tuning **PI** controller is an adaptive **PI** controller which its coefficients are tuned using a neural network[11]. In the neural network, weights are updated at each sampling time by means of the error between the desired output and the actual output of the system. Because the growth of biomass is an exponential function of μ, so with handling μ as the output of the system the biomass concentration is implicitly controlled [4]. As already illustrated, the μ_{ref} considered here is that of figure(2). Schematic of the Neural network based self-tuning **PI** controller is explained in figure(2):

In this scheme a multilayer neural network is utilized which has seven neurons in the hidden layer and two neurons in the output layer. Two outputs of NN are the k_i and k_p coefficients of **PI** controller. The activation function of neurons in the hidden layer is chosen to be a sigmoid function while a linear function is used in the output layer because in contrast with the sigmoid function there is no limitation in the range of the linear function and so there won't be any limitation in the output range. Training of the network is done on-line and there is no need to any off-line initial learning on weights. The learning rule applied in this neural network is the famous back-propagation rule utilized in multilayer neural networks.

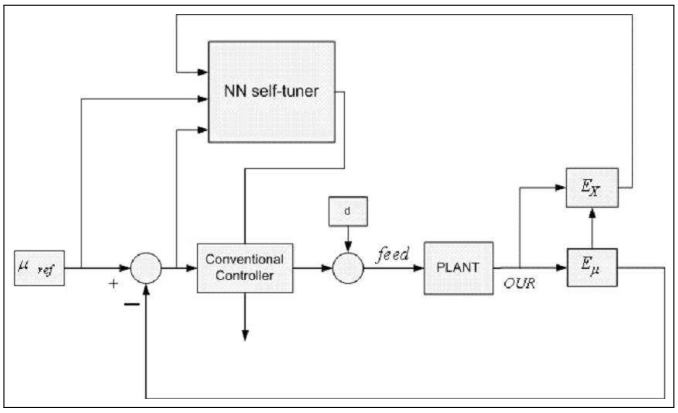


Figure 2: Schematic of the NN based self-tuning **PI** controller

For tracking of the desired specific growth rate, μ_{ref}, at first a fixed gain **PI** controller was designed. Due to nonlinear characteristics of process, controller's gains were tuned manually. The concluded controller could track μ_{ref} in a rational way. In addition, it had a good robustness in front of input disturbances and rejected them. This controller also could resist in front of parameter variations such as S_{in}, $Y_{X/S}$ and tracked the setpoint. But this controller had a big disadvantage that with changing the reference specific growth rate to a constant setpoint, it couldn't track new setpoint. This phenomenon is natural because each setpoint needs its coefficient tuning. But an adaptive controller can easily track feasible setpoints by tuning the coefficients on-line. So by applying a neural network based self-tuning **PI** controller this fault of the conventional **PI** was surmounted meanwhile all advantages of the fixed gain **PI** controller such as robustness in front of parameter variations and disturbance cancelation was maintained. Figure(3) shows specific growth rate tracking in fed-batch stage using NN based self-tuning **PI** controller. One can see that after depletion of substrate which causes sudden decrement of μ, the fed-batch stage launches as illustrated before.

In figure(3) it can be seen that the controller appropriately tracks specific growth rate setpoint in the fed-batch stage. So with changing the specific growth rate setpoint to the constant value of $\mu_{ref} = 0.3$ it is clear that adaptive controller could easily track new setpoint.

Like the conventional **PI**, NN adaptive controller could also reject big disturbances and parameter variations. Figure(4) shows a disturbance added to the input of process from time $t = 11[h]$ to time $t = 14[h]$ and it is obvious that the controller could cancel it appropriately. In figure(4) output of the controller (i.e. the feed entered to system)

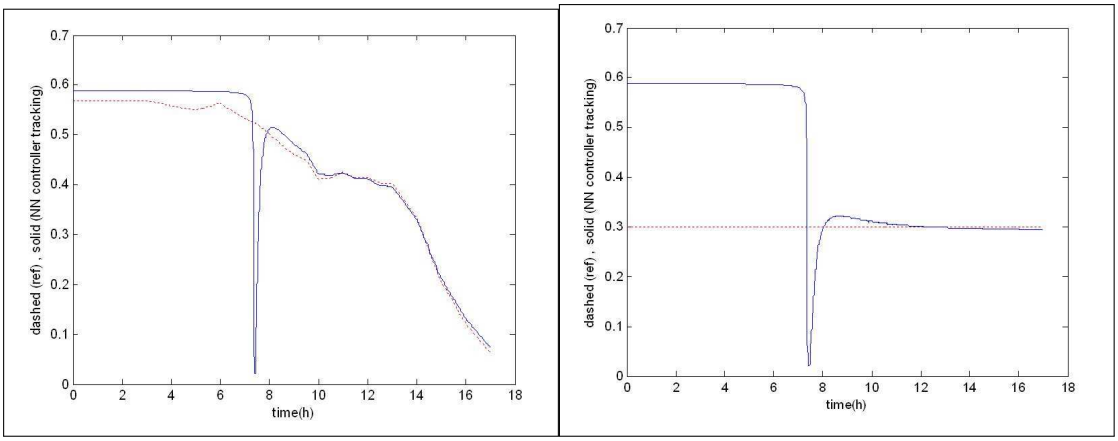


Figure 3: μ_{ref} and corresponding μ gained from NN based self-tuning **PI** controller

in the case of A) in presence of disturbance $d = 25\lfloor cm^3h^{-1}\rfloor$, have been shown and it is clear that if output of the controller in presence of disturbance is aggregated with disturbance, the feed attained is equal to the output of controller when there is no disturbance. This means that controller could truly cancel the effect of disturbance.

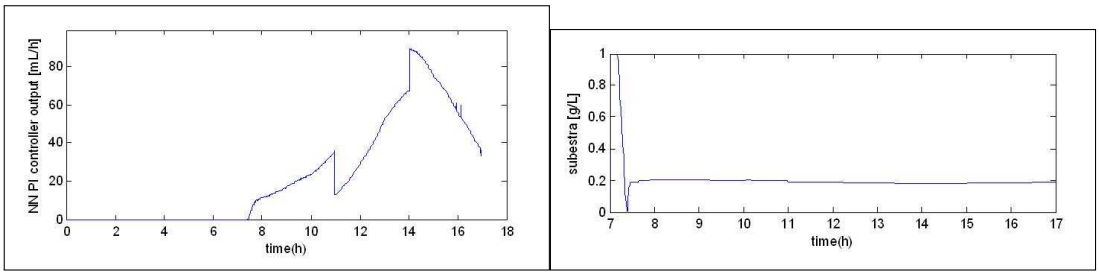


Figure 4: A) controller output in presence of disturbance B)output of process ,μ , in presence of disturbance

The only deficiency of NN based self-tuning PI controller is the lack of convergence proof so that there is no convergence guarantee to setpoints.

2.2 Optimal control of fed-batch process

In order to attain maximum performance of the system, designer should keep glucose concentration in the critical value S^{crit} to achieve maximum biomass growth meanwhile keeping acetate concentration minimum. According to equation(1) :

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu X + \frac{F}{V}(S_{in} - S) \tag{3}$$

Based on the assumption that the (kinetics in the) model is perfectly known, the minimum principle of Pontryagin [12] provides the optimal feed rate F^*, which maximizes a prespecified control objective. If the control objective is to maximize the final amount of biomass the optimal control law would be the following [5, 6]:

$$F^* = \frac{\frac{\mu}{Y_{S/X}}X}{S_{in} - S^*}V \tag{4}$$

This control law keeps the substrate concentration constant ($\frac{dS}{dt} \approx 0$) at a prespecified setpoint S^* and equals the exponential feed rate. If controller (equation(4)) is implemented in open-loop, process disturbances whether measurement or modeling errors can't be compensated. Therefore some mechanism must be incorporated in the control law (equation(4)) to control the tracking error in the presence of disturbances. This is done e.g. by adding a feedback linearizing term or sliding mode controller.

2.3 Feedback linearizing control

The following closed-loop dynamics for the substrate concentration ,S, is imposed to guarantee the convergence of the controller to the desired setpoint S^* (with τ_S $[1/h]$ a strictly positive convergence rate factor):

$$\frac{d(S-S^*)}{dt} = -\tau_S(S-S^*) \tag{5}$$

The combination of the closed-loop dynamics (equation(5)) with the mass balance equation for the substrate concentration (3) results in:

$$F = \frac{\frac{\mu}{Y_{S/X}}X}{S_{in}-S^*}V - \tau_S\frac{S-S^*}{S_{in}-S^*}V \tag{6}$$

The linearizing control law (equation(6)) can be interpreted as the feedforward optimal control (equation(4) first term) plus feedback action (second term).

According to equation(2) for kinetics, if the effect of acetate is neglected and just the base Monod model is considered, specific growth rate μ will be a monotonic function of substrate S and there will be a unique value for μ with each value of substrate.

In this part the feedback linearizing controller is examined with critical growth rate of $\mu^{crir} = 0.52$ which is equivalent to substrate concentration $S^{crit} = 0.20$. The value of the considered τ_S is equal to $50[1/h]$ which was acquired by trial and error. Figure(5) shows tracking of $S^* = S^{crit} = 0.20$ by means of feedback linearizing controller:

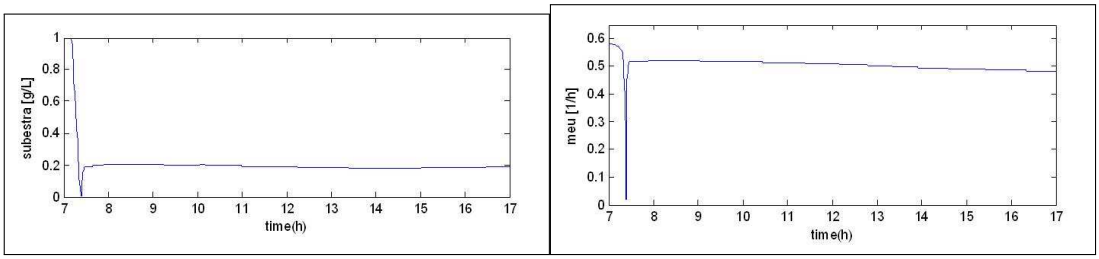


Figure 5: tracking of substrate setpoint $S^* = .2$ by means of feedback linearizing controller A) substrate concentration in fed-batch stage C) μ in fed-batch stage

This controller has good resistance against some disturbances while it is very sensitive against other parameters which are directly multiplied in the manipulated input of equation(6) such as S_in and volume of liquid phase(V). Figure(6) shows tracking of substrate setpoint $S^* = 0.20$ and its equivalent $\mu^* = 0.52$ while changing S_{in} from $600[g/L]$ to $500[g/L]$ at time $t = 10[h]$.

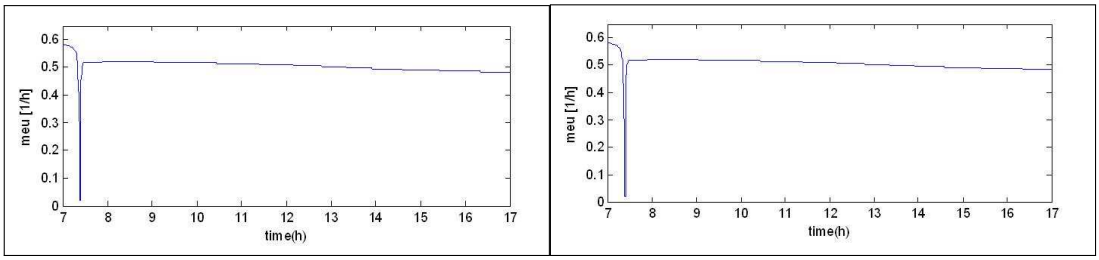


Figure 6: tracking of substrate setpoint $S^* = 0.20$ while changing S_{in} from $600[g/L]$ to $500[g/L]$ at time $t = 10[h]$.

3 Conclusion

Comparison of above feeding control strategies show that feedback linearizing controller has good performance and can be used as a practical method which its convergence is ensured. But it has the disadvantage that it is

designed for fixed setpoints and can't be used for tracking empirical setpoints such as figure(2), the problem that NN based self-tuning **PI** controller could easily obviate but it doesn't have any convergence guarantee.

In addition, comparison of these controllers in tracking fixed setpoints show that feedback linearizing controller performs better because NN based self-tuning **PI** controller needs more time to train its coefficients and converge to the setpoint. But NN based self-tuning **PI** controller is more robust against parameter variations such as S_{in}, V and also can be used for tracking empirical setpoints.

References

[1] M. Åkesson, "Probing Control of Glucose Feeding in Escherichia coli Cultivation" , *Phd thesis*. Lund. 1999.

[2] V. Babaeipour, S. A. Shojaosadati, S. M. Robatjazi, R. Khalilzadeh, N. Maghsoudi "Over-production of human *interferon-*γ by HCDC of recombinant *Escherichia coli*" Process Biochemistry, 42, pp. 112-117, 2007.

[3] V. Babaeipour, S. A. Shojaosadati, R. Khalilzadeh, N. Maghsoudi, F. Tabandeh "A proposed feeding strategy for over-production of recombinant proteins by *E.coli" Biotechnology and Applied Biochemistry*, Article in Press, 2007.

[4] L. de Maré, P. Hagander "PARAMETER ESTIMATION OF A MODEL DESCRIBING THE OXYGEN DYNAMICS IN A FED-BATCH E.COLI CULTIVATION." *Biotechnology Letters* **27:14** , pp. 983-990, 2006.

[5] J. Lee, S. Y. Lee, S. Park, A. P. J. Middelberg, "Control of fed-batch fermentations" *Biotechnology Advances*, 17, pp. 29-48, 1997.

[6] N. Nedjah, L. de Macedo Mourelle, *Swarm Intelligent Systems*, Springer-Verlag Berlin Heidelberg, pp. 3-10 , 2006.

[7] R. Neeleman, "BIOMASS PERFORMANCE Monitoring and Control in Bio-pharmaceutical Production" , *Phd thesis*. van Wageningen Universiteit. 2002.

[8] F. Renard, A. Vande Wouwer, S. Valentinotti "A practical robust control scheme for yeast fed-batch cultures- An experimental validation" *J of Process Control*, No. 16, pp. 855-864, 2006.

[9] O. Roeva, St Tzonkov "Modeling of *Escherichia coli* Cultivations: Acetate Inhibition in a Fed-batch Culture," *Bioautomation*, 4, pp. 1-11, 2006.

[10] I. Y. Smets, G. Bastin , J. Van Impe "Feedback Stabilization of Fed-Batch Bioreactors: Non-Monotonic Growth Kinetics" *Biotechnology*. *Prog*, Vol. 18, No. 5, pp. 0005-0014, 2002.

[11] M. Teshnehlab, K. Watanabe , *Intelligent Control Based on Flexible Neural Networks* , KLUWER ACADEMIC PUBLISHERS , pp. 42-56 , 85-104 , 1999.

[12] J. F. Van Impe, G. Bastin "Optimal adaptive control of fed-batch fermentation processes." *Control Eng*. Practice, 3, 939-954, 1995.

Saleh Mohseni
Islamic Azad University, Ayatollah Amoli Branch; Amol, Iran
E-mail: s_saleh_mohseni@yahoo.com

Ahmad Reza Vali
AmirKabir University of technology, Tehran, Iran
E-mail: ar.vali@gmail.com

Valiollah Babaeipour
Tarbiat Modares University, Tehran, Iran
E-mail: vbabai@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 402-407

An Agent-holon Oriented Methodology to Build Complex Software Systems

Gabriela Moise

Abstract: In real life, there are a lot of complex processes which have to be automated in order to be more manageable and better controlled. The programmers have to build complex software applications which are to serve the complex activities of an organization. The agent-oriented paradigm has developed for the last decade and new techniques of software engineering have been issued, such as the agent-oriented software engineering. In this paper the author presents a study of the methodologies defined to the analysis and design agent-based system and proposes a methodology based on the concepts of agent and holon to build a complex system. One concept is the concept of "holon", considering that any organization has to be viewed according to a holonic structure and the other concept is the concept of "agent", a software entity with some special properties (autonomy, reactivity, social abilities, learning capacity, goal orientation, etc.). The proposed methodology fits the construction of the complex software system with fine granularity.
Keywords: agent software, holon, software engineering

1 Introduction

A lot of complex and different activities carry on within an organization. The nature of business processes and the IT systems are getting more and more complex and sophisticated. To describe the complexity of a system, it is necessary to understand how the relationships between components determine the behaviour of the system, how the behaviour of each part of the system contributes to the whole behaviour of the system and how the system interacts with the environment and causes changes in the environment.

The science of complex systems has been acknowledged as a new science that chips away all boundaries of the traditional disciplines: engineering, management, medicine, philosophy, social sciences, ecology, education, environment, and so forth. It is really hard to provide a measure of the system's complexity. Some attempts were done in order to measure the complexity of a system: computational complexity (time measures or number of steps); information measuring (applying the algorithmic information theory); using degrees of comparison between things, in which at least one's complexity can be measured; using the granularity of the system's entities [8].

Summing up, "the complexity of the system is the amount of information necessary to describe it" [2]. If one takes into consideration the above mentioned facts, it is imperative to develop a methodology to build complex system. The most complex systems are based on many agents and a special case of complex systems are the complex adaptive systems. The researches agree on the fact that complex systems tend to rise into more complex systems. This aspect is a challenge for a software developer. According to [1], a systems' methodology requires: "(a) the analysis of systems and systems problems, problems concerned with the systemic/relational aspects of complex systems; (b) the design, development, implementation, and evaluation of complex systems; and (c) the management of systems and the management of change in systems." Another aspect is that even if the structure of the components of the system changes, the behaviour of the system has to remain the same or has to be improved. So, how to develop a complex system is a challenge and in this paper, the author presents a methodology based on two concepts - i.e. those of agent and of holon - to develop a complex system. The remaining part of the paper is organized as follows: in section 2, there are presented some methodologies to develop agent-oriented systems, in section 3, the methodology itself and a case study and in section 4, the Summary and the Conclusions, the work to be done in the field of the complex system.

2 Methodologies to Develop Complex Systems

The agent-oriented paradigm has developed and it is used to build complex software systems from different areas: medical applications, industrial applications, such as manufacturing, air traffic control, process control, telecommunications, transportation systems, commercial applications, such as electronic commerce, business process management applications, entertainment applications [4].

Despite the great number of agent-based applications, there is a consensus on the fact that there is a crucial lack concerning specification and development methodologies to build agent-based systems [3].

Many approaches take the OO techniques and methodologies and extend or adapt them to domain applications in order to design the agent systems. While the mental state of agents can not be modelled using OO techniques, other approaches extend the Knowledge Engineering techniques. Most of them restrict to the phases of analysis and design and they do not provide a tool to implement an agent-based system. Here are some methodologies to analyze and design agent-based systems:

- Gaia (Wooldridge, Jennings and Kinny, 1999);
- Prometheus (Lin Padgham and Michael Winikoff, 2002);
- Multiagent Systems Engineering (MaSE) (DeLoach, 1999);
- KaOS (Bradshaw et. al, 1997);
- Agent-Oriented Analysis and Design (AOAD); (Burmeister, 1996)
- AOM (Shoham, 1993).

In the following paragraphs, the author briefly presents two methodologies, which she considers useful in developing MAS. In the methodology (GAIA), presented in [11], the process of building a MAS is viewed as a process of organizational design. The most abstract entity in the concepts' hierarchy of GAIA is the concept of system. The first step of GAIA methodology is to identify the main roles of the system. Different roles in an organization interact, using a set of protocols in order to achieve their own goals and to contribute to the objectives of the organization. Responsibilities are the key attribute of a role in that they determine its functionality. A role has rights, which were called permissions. The organizational model is defined through two models: the roles models and the interaction ones. The design process generates three models: the agent model, the services model and the acquaintance model (figure no.1).

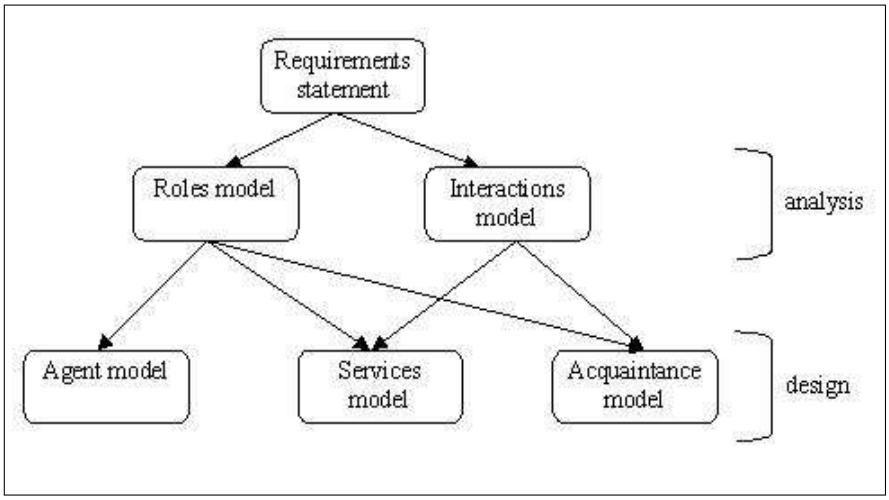


Figure 1: Relationships between Gaia methodology models (presented in [11])

The Prometheus methodology defines a generic modelling language that fits any MAS architecture (even its authors consider that this methodology is not so useful for non BDI agents) and any implementation environment. This methodology consists of three phases: the system specification phase, the architectural design phase and the detailed design phase. The roles of the system care, called functionalities, are identified in the system specification phase along with inputs, outputs and shared data sources. In the architectural phase, there are identified the types of agents, share data objects and there are provided the system overview diagrams and interaction diagrams. In the detailed design phase, there is the description of every agent's internal structure and of the way it will accomplish its tasks within the overall system. The author of the present paper considers this methodology more practical than others and a valuable tool to develop MAS using Prometheus at [12]. More details about Prometheus methodology can be found in [9] and [10] . There is an agreement on the fact that a Unified Agent-oriented Modelling Language

(UAML) is necessary, as a methodology that has to fully support the requirements, analysis and design phases, and the tools used for describing the complexity of multi-agent systems.

3 A Methodology Used for Building Complex System

A new methodology has been considered starting from the point of view that complex systems are composed of software entities that are not really intelligent agents. Most of the applications use the attribute ”intelligent”, but they are not so intelligent, as one can say that there is no brain. This ambiguity is also sustained by the fact that they have not reached an agreement yet about the definition of the intelligent agent term. So, the author considers that a complex system can be built using software entities with the following possible attributes:

1. Autonomy;
2. Mobility;
3. Veracity;
4. Having reason;
5. (Having) character;
6. Flexibility;
7. Robustness;
8. Goal oriented;
9. Learning capacity;
10. Good will.

The methodology proposed in the paper is based on two concepts: ”agent”, already presented, and ”holon”. The concept of ”holon” was proposed by Koestler (1967) in his book *The Ghost in the Machine*.[5] The word ”holon” is a combination of two words: the Greek ’holos” meaning ”whole” and the suffix ”-on” meaning ”a particle” or ”a part”. Koestler asserted that parts and wholes in an absolute sense do not exist in the domain of life. So a holon is ”an identifiable part of a system that has a unique identity, yet is made up of sub-ordinate parts and in turn is part of a larger whole” [5]. A holon can be part of another holon, or a holon can be broken into several other holons, which in turn can be broken into holons. The minimum attributes of a holon are: autonomy and cooperation. These entities can be seen in a holarchical relationship with each other. Koestler called the systems of such entities Open Hierarchical Systems. There is a similarity between the concept of the holon and the concept of agent, and the differences between these concepts tend to be reduced. FIPA and HMS consortiums introduced guidelines and specifications that support the holonic requirements. A ”holarchy” is a system of holons that can cooperate to achieve a goal or objective. The benefits of the holonic organisation are: stability, flexibility, adaptability. In this paper two terms are used: simply software entity to denote an entity without any attributes; abstract agent holarchy to denote a holarchy of agents or simply an agent. An agent is a software entity with some attributes from the list from above and uses simply software entities.

The way in which the models are defined is inspired from the work of building a complex system called Production Information Management System, implemented in a packing manufacture [7]. This methodology consists of the following phases: the system goals analysis, the architectural design phase and the detailed design phase and Prometheus Design Tool is used [12]. In the first phase a goals hierarchy is realized, starting from a general objective and ending with the most atomized goals. In figure no. 2 is presented a generic diagram of goals. For each goal from the level no. 1 it is designed and implemented an abstract agent holarchy .

Unlike the first stage, where a top-down approach is used, now, in the second phase they use a down-top approach. So, a set of atom goals is attained.

In the second phase, the architectural design one, simple software entities and abstract agent holarchy to achieve the goals are identified, as presented in figure no. 3. The atom goals are grouped and a group of atom goals are related to an agent. So, an agent gets a set of goals.

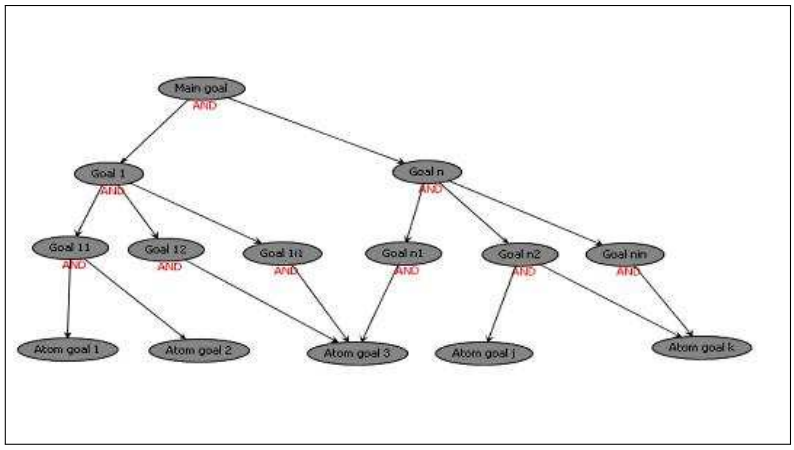


Figure 2: Goals diagram

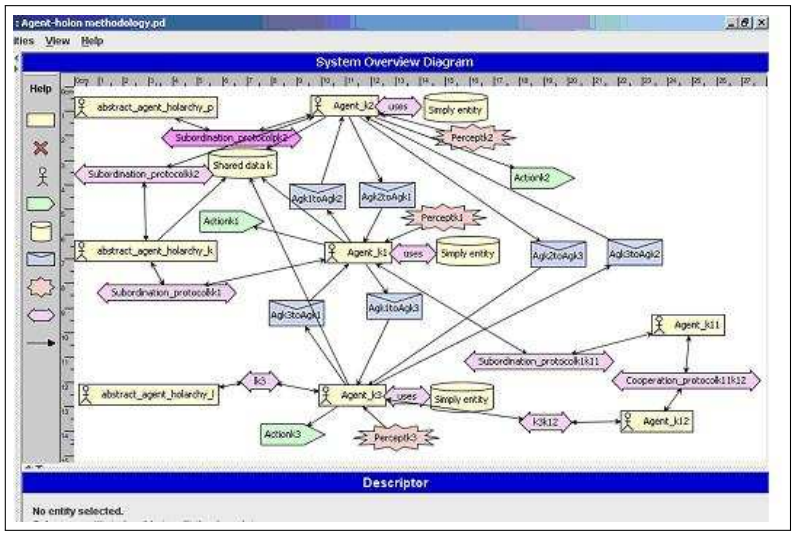


Figure 3: System overview diagram

The system's diagram shows all components of the system and all dependencies between them. This diagram is useful to design all details and to make a planning of application's development. Also, we have to remind that we can make an estimation of the costs involved to build, implement and maintenance of the software application.

At third phase, detailed design phase the internal structure of each agent and simply software entities, subordination protocols and co-operation protocols are designed (figure no. 4).

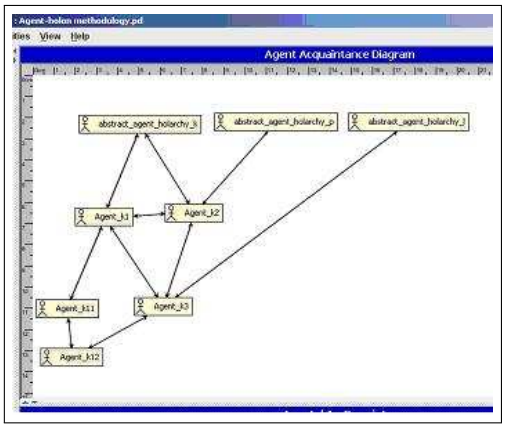


Figure 4: Detailed design

They are two types of protocols: co-operation protocols and subordination protocols. The agents communicate between them using these protocols and messages. (figure no. 5)

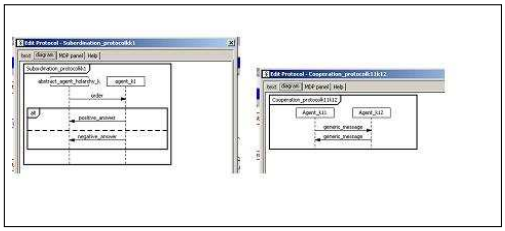


Figure 5: Protocol design

The internal structure of each agent contains the following possible of entities: 1. Action , 2. Capability, 3. Data, 4. Message, 5. Plan, 6. Percept. A case study of using the agent-holon oriented methodology is presented in figure no. 6.

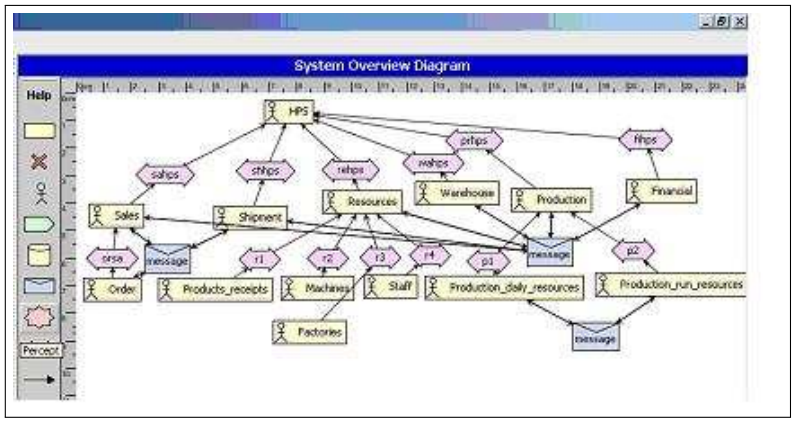


Figure 6: Production Management Diagram

The complexity of the system is growing in time. We start from building simple agents, software entities, agents holarchy and we finish with a linking procedure to generate the whole system. We add more holarchy at

system using protocols therefore the system become more complex and functional. The one of the most important characteristic feature of this system consists in that fact that even if one of the component changes the functionality of the system is not affected.

4 Summary and Conclusions

The originality of this methodology consists in the fact that an objective is decomposed in a tree of goals, each branch being solved by building an abstract agent holarchy, be it a holarchy of agents or simply an agent. In the first phase of the methodology they use a top-down approach. After that each holarchy is developed, encoded and implemented. Finally the holarchies are linked and the application is generated. Each holarchy of agents can use simple software entities and solve a set of goal. The complexity of the system is growing in the phase of implementation, through building each holarchy and during run time, while new data acquisition is taking place, or plans are changed. This methodology has the following advantages: it enables one to atomize the problems and so simple problems can be solved more easily, generates portable abstract agent holarchy, in the sense of using them in other systems, makes the maintenance of the complex system easier and enables control application.

References

[1] B. H. Banathy, P. M. Jenlink, "Systems Inquiry and Its Application in Education", *In D. H. Jonassen (ed), Handbook of Research for Educational Communication and Technology, New York: MacMillan Library Reference*, 1996.

[2] Y. Bar-Yam , "Dynamics of Complex Systems (Studies in Nonlinearity)", *Advanced Book Program*, Westview Press, 2003.

[3] V. Hilaire, A. Koukam, P. Gruer, Jean-Pierre Müller, "Formal Specification and Prototyping of Multi-Agent Systems", *Lecture Notes in Computer Science* 1972:114-227, 2001.

[4] N. R. Jennings, K. Sycara, M. Wooldridge, "A Roadmap of Agent Research and Development", *Autonomous Agents and Multi-Agent Systems*, Kluwer Academic Publishers, Boston, 1, 7-38 ,1998.

[5] A. Koestler, *The Ghost in the Machine,* Hutchinson & Co, 1967.

[6] V. Marik, M. Pechoucek, P. Vrba and V. Hrdonka, FIPA Standards and Holonic Manufacturing, *Agent-based Manufacturing: Advances in the Holonic Approach,* pages 89-121. Springer Verlag, 2003.

[7] G. Moise, Production Information Management System, Information & Knowledge Age, *The Proceedings Of The Seventh International Conference Of Informatics In Economy,* ASE Bucuresti, 2005.

[8] Murray Gell-Mann, "What is complexity?", John Wiley and Sons, Inc.: Complexity, Vol. 1, no. 1,1995.

[9] L. Padgham and M. Winikoff , "Developing Intelligent Agent Systems: A Practical Guide", RMIT University, Melbourne, AUSTRALIA, Published by John Wiley and Sons, 2004.

[10] L. Padgham and M. Winikoff. "Prometheus: A Methodology for Developing Intelligent Agents".*Proceedings of the Third International Workshop on Agent-Oriented Software Engineering*, at AAMAS'02, http://www.cs.rmit.edu.au/agents/Papers/aamas02-aose-ws.pdf, 2002.

[11] M. Wooldridge, N. R. Jennings, and D. Kinny, "A Methodology for Agent-Oriented Analysis and Design", *Proceedings of the Third International Conference on Autonomous Agents (Agents'99)*, 1999.

[12] http://www.cs.rmit.edu.au/agents/pdt/.

Gabriela Moise
Petroleum-Gas University of Ploiesti
Computer Science Department
no. 39 Bd. Bucuresti, Ploiesti, Romania
E-mail: gmoise@upg-ploiesti.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 408-412

WebAgeing - A Flexible System for Personalized Accessing of Services for Ageing Population

Maria Moise, Victor Popa, Marilena Zingale, Liliana Constantinescu, Alexandru Pirjan

Abstract: Accessing public services for different domains including: health, social programs, taxes etc, is a bureaucratic challenge for all citizens, but the ageing population is highly affected. To manage this state of facts, the government agencies, non-government and other organizations offer online form for accessing services dedicated to this segment of the population. Sadly, this effort has had minor effects regarding the improvement of the existing situation. In this context, the paper presents a solution to design, conceive and integrate a "middle ware" type infrastructure based on semantic technology, capable of providing composed personalized services for the ageing population according to their profile. The proposed approach intends to integrate independent, heterogenic, geographically distributed services, offered by service providers (government agencies, non-government, other organizations) elderly citizens, using a registering mechanism for services with the semantic communities, based on matching syntactic and semantic algorithms, and also compose providers' services in a personalized manner, in the purpose of executing demands addressed to the ageing population. Communities' network will be completed by a guide in order to accomplish and to exploit the type "one-stop" platforms at all levels (national, regional and local) in concordance with EIF (European Interoperability Framework). The ideas of this paper come from a research project, financed by a National Program - PNCDI II during 2007 - 2010.
Keywords: Composed personalized services, syntactic matching algorithms, user profile.

1 Introduction

The main objective of the flexible system, entitled WebAgeing is building an infrastructure based on WEB services and semantic technologies with the purpose of improving the access to services and programmes destined for the ageing population.

The objectives refer to:

- demands regarding access to services and programs destined for the elderly;
- service modeling methodology, support instruments, referring to the modeling of services using WSDL standard extended with semantic and quality attributes;
- the ontology of modeling the community, support instruments, referring to the modeling of the communities using syntactic, semantic and operational attributes;
- the methodology of web service registering with semantic communities;
- specifying the users demand methodology, support instruments;
- composing and execution of services methodology, support instruments;
- algorithms for personalized generation of composed services, execution of composed generated services;
- components for creating the complete WebAgeing platform, allowing community providers to define their communities by using Community ontology, service providers to register services for semantic communities and the ageing population to specify its demands. A consistent implementation of the prototype for a complete WebAgeing platform will be conceived by exploring as much as it ca possibly be explored the existing technologies.

In order to achieve these objectives WebAgeing takes into consideration the European approach regarding service accessibility and initiatives concerning semantic Web Services. Regarding services' accessibility, WebAgeing is based on 'Web Accessibility Initiative' (WAI) created by W3C group that offers a guide concerning the projection and structure of Web services. Regarding the semantic initiatives for the modelling of Semantic Web Services (SWS) there are two types:

- initiatives that require anthologies defining for the representation of any aspect of SWS (for example OWL-S, WSMO);

- initiatives that encourage the extent of the actual SWS, with open possibilities to reason over SWS definitions, in the purpose of extracting service capabilities, in order to match these capabilities with the ones demanded by the clients (for example METEOR-S, WebDG, WSMX).

The first class of initiatives defines concepts capable of semantic descriptions of all Web service aspects. These aspects normally include the functionality offered by the service (characterized by preconditions and effects), interface and non-functional characteristics of services etc. The second initiative uses additionally, a collection of intelligent instruments specialized in defining, publishing, discovering and appealing to Web services.

In actual systems there are a tremendous number of possible combinations during the generation of composed services. By using WebAging approach, only services with relevance to the related communities are combined in a composed service, for example you can combine registered services with Health community and registered services with Disability community.

The matching algorithms used in WebAgeing consume significant less time then actual systems. They consume execution time only at the moment of service registering with the semantic communities, at the time when the degree of matching of generic operations with concrete operations is calculated. At the time of emitting of the demands by the elderly, WebAgeing generates composed personalized services as a response to those demands, using the matching degree already calculated, without executing the matching algorithms again and so, the user waits less to obtain the answer.

Composing services implies knowing the rules and regulations in the service and program accessing domain for the ageing population. WebAgeing doesn't demand this knowledge from the user (senior citizens, public officers) because the composition is created in a dynamic way by the system and the knowledge in the domain are furnished by semantic communities.

The discrepancy on in-out messages of service components from inside composed services. WebAgeing establishes the correspondence between in-out parameters of generic operations and in-out parameters of service operations, from the moment of service registering with semantic communities. The correspondence is based on the syntactic and semantic attributes of message parameters.

WebAgeing uses the user profile to describe security and confidentiality policies. WebAgeing responds differently to each demand, addressed by users that have different profiles. The rules of eligibility attached to generic operations take in consideration the values detained by user profiles.

2 Description of the WebAgeing system

2.1 WebAgeing's approach

WebAgeing approach, being part of the second type of initiative, takes in consideration the following technologies and standards: WSDL, WSDL-S, UDDI, RDF-S, OWL, OWL-S, SOAP, HTTP, RMI.

The following European projects address certain aspects relevant to WebAgeing approach:

- *SENIORWATH* addresses the need for a better understanding and monitoring of web services for the ageing population;

- *HealthService24* is a eTEN project that validates a new platform for the continuous elderly monitoring. Thus, the senior citizens are equipped with sensors, driven by a mobile phone. The measurements are sent wireless to a medical centre where the data is analysed immediately and the personalised feed-back is sent to the patient in real time;

- *SENIORITY*, a eTEN project that uses ICT to furnish quality models for the European sector for care for the elderly;

- *Onto Gov* (Ontology-enabled E-Government Service Configuration) (http://www.ontogov.com/), a deploying IST project whose objectives are the development, testing and validating a semantic-enriched platform (using anthologies) that will ease the consistent composition, reconfiguration and evolution of government services;

- *Semantic Gov* is a project that helps the analysis, display, implementation and evolution of an intelligent and integrated platform for furnishing services in public administration. The project is based on the paradigm of Architecture oriented Services (SOA), implemented through the Semantic Services technology.

Other projects taken in consideration for the approach of WebAgeing are: *WebDG* that concentrates on the composition of Web services and on keeping the data confidentiality, *METEOR-S* that furnishes the semantics at the data level, functional, non-functional and executive.

WebAgeing approach exploits the results obtained from projects mentioned above and adds to their results, mechanisms for:

- organizing services destined for the elderly in communities based on semantic domains, communities that are defined by a set of semantic attributes that identify the community (community category, community synonyms, and community specializations). A collection of generic operations where each generic operation is identified, also, through semantic attributes (operation category, purpose of the operation, operation synonyms, and operation specializations) include in-out parameters, rules of the operation eligibility etc. For the loading of semantic attributes values different standards are used, like: NAICS (North American Industry Classification System), UNSPPSC (Universal Standard Products and Services Classification), Rossetta Net, cXML, EDI etc;

- service registering with semantic communities, using matching algorithms between generic operations of the communities and concrete operations of Web services;

- personalized composition of services based on rules and regulations from the domain of accessing services and on user profile. The rules and regulations for accessing services by the ageing population will be implemented using relationships between generic operations (pre-operations, post-operation) and eligibility rules attached to generic operations;

- the interrogation of communities‘ collection as a data base. The user has the possibility of consulting the communities in the purpose of selecting generic operations to be included in the users‘ demands;

- ensuring information security and confidentiality for Web services, using credentials and data filters. The security and confidentiality policies can be found and recorded with the user profiles.

2.2 WebAgeing's activities

In order to accomplish the above mentioned objectives there have been define the following activities:

1. *demands analysis*
   Demands analysis consists on the definition of the system for personalized access of the elderly to specific services and programs in terms of software and architecture design.

2. *specifying the community modeling method, Web services, recording of services, semantic communities*
   This activity implies defining the domain, generic operations, relations between operations, eligibility rules attached to generic operations, messages, parameters, quality attributes, service recording method with the communities. Communities are created by community providers (alliances of agencies, non-profit organizations and other organizations that have in common an interest domain), using Community ontology as a creation template. The generic operations include, along the syntax, their semantic description, messages and generic operation parameters that include, along syntactic description also a semantic description. WebAgeing adopts UML activity diagrams to graphically represent 'pre-operation' and 'post-operation'. Eligibility rules attached to the operations use logic conditions that include variables assignment that is conceived in the user's profile. Web services are defined by using syntactic structures from WSDL to which are attached semantic attributes for identifying the service (category), semantic attributes for identifying the operations (purpose), semantic attributes attached to messages and parameters, quality attributes (security, price, answering time etc) attached to services. Communities are registered in the Community Registry using syntactic attributes (communities ID) and semantic attributes (Category, Synonyms, Specializations) in the purpose of efficient discovery at the time of the development of composed personalized services, used to provide answers to users demands. The related communities are interconnected. Web Services described using

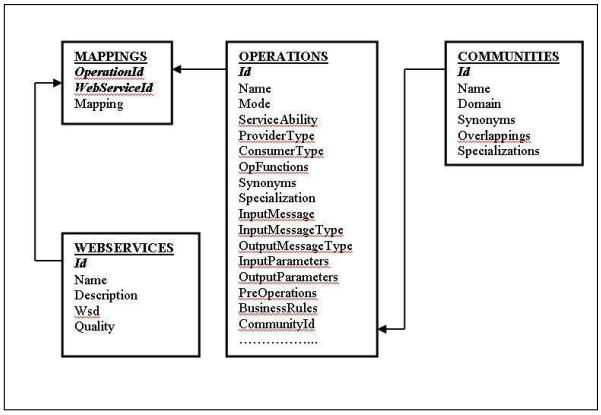


Figure 1: Community Ontholohy

extended WSDL is registered with semantic communities. The fig.1 illustrates the relationships in-between Communities, Generic Operations and Web Services.

Registering Web services with semantic communities implies building mappings between generic operations of the communities and concrete service operations, using matching algorithms between community attributes and Web service attributes mappings between in-out parameters of the generic operations and in-out parameters of the service an operation using matching algorithms between parameters attributes. Matching algorithms calculate the degree of matching that they memorize in the Mapping object. A Web service can register with many communities if it implements generic operations defined in those communities (fig.2).

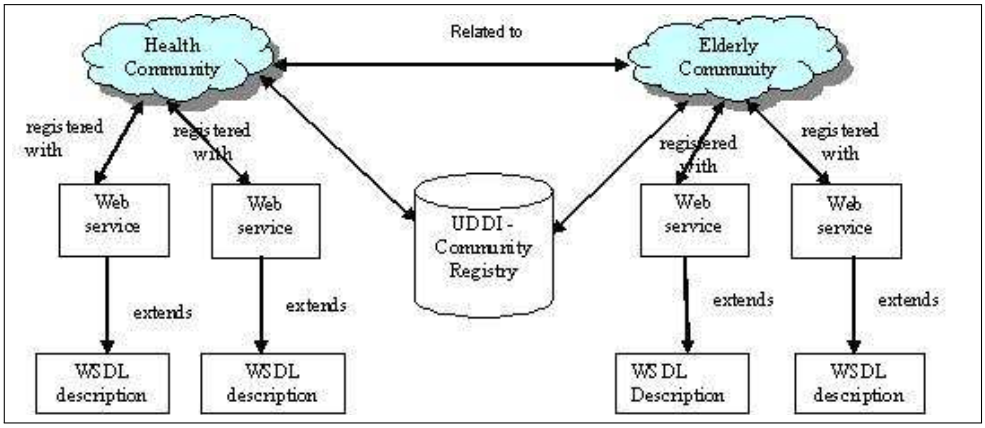


Figure 2: Registering Web services with semantic communities

3. *specification of the users demands, the methodology for personalized composition of services and methodology for execution of services*
   This activity consists of defining specification formalism of the user demands and the methodology for personalized generation of composed Web services; definition of the machinery for the execution of composed services. The senior citizens specify the demands by selecting generic operations relevant for their purpose, from the semantic communities.

4. *selection of the technology*
   This activity will detail the specified architecture, identifying relevant software components. Instruments and components necessary to the WebAgeing system will include: communities management registry instruments, communities management instruments, instruments for the Web service registering with semantic communities, instruments for the demands management and user profiles, instruments for building composed personalized services, instruments for the execution and monitoring of composed services.

5. *development and validation of the system*
   System development will be accomplished through the following tasks: development/customization of instruments for the management of communities registry, development/customization of instruments for communities management, development/customization of instruments for the management of Web service registering with interest communities, development/customization of instruments for demands and profile user management, development/customization of instruments for building composed personalized services, instruments for execution and monitoring of composed services.

3 Conclusions

For the agencies and organizations that offer services for the ageing population, WebAging is motivated because of the necessity of organizing this services in the semantic communities. For the public workers that deal with the elderly, WebAgeing is motivated through the necessity of solving demands of programs and services, addressed by the ageing population. For the elderly that prefer demanding programs and services from the personal computer, WebAgeing is motivated through the necessity of building personalized composed services, which through out their execution to offer answers to the addressed demands. For the communities‘ providers, WebAgeing is motivated through the necessity of defining an ontology that world serve as a template for the definition of the communities.

Main impact through WebAgeing system exploitation represents:

1. encouraging the development of a technological platform for the improvement of access to services and programs destined for the ageing population;
2. collaboration stimulation in an organizational framework of agencies and organizations that provide services and programs for the elderly;
3. diminishing the difference in the life level between the ageing population in Romania and ageing population abroad;
4. eliminating the bureaucratic practices existent in the administrative system, practices that create great problems to every citizen, especially to senior citizens.

References

[1] A. Arkin, *Business Process Modeling Language*, San Mateo, CA: BPMI. org, Proposed Final Draft, 2002.

[2] A. Banerji, C. Bartolini, D. Beringer, V. Chopella, K. Govindarajan, A. Karp, H. Kuno, M. Lemon, G. Pogossiants, S. Sharma, and S. Williams, *Web Services Conversation Language (WSCL) 1.0.*, World Wide Web Consortium, W3C Note., 2002.
BPMI.org. *Business Process Modeling Language (BPML)*, Alameda (CA): Business Process Management Initiative, Working Draft 0.4., 2001.

[3] M. Moise, V. Popa, L. Constantinescu, *A generic solution for business process management across public institution boundaries*, in Proc of Information Systems and Operations Management Workshop, Bucharest, pp. 65-74., 2006.

Maria Moise, Alexandru Pirjan
Romanian American University, Romania
E-mail: {maria.moise,alex}@rau.ro

Victor Popa, Liliana Constantinescu
National Institute for Research and Development in Informatics
E-mail: {vpopa,}@ici.ro

Marilena Zingale
USA
E-mail: zingale@un.org

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 413-416

Advanced Modelling of Tutor Intelligent Systems for Distance Learning Applications

Ioana Moisil, Iulian Pah, Dana Simian

Abstract: In this paper we are presenting a model for an adaptive multi-agent system - ACTIVITIES system - for dynamic routing of the learning activities' tasks of a learning environment, based on the adaptive wasp colonies behaviour. The presented model allows the assignment of activities taking into account the qualifications of students, their experience and the complexity of tasks already performed. The system is changing dynamically, because both the type of activities and the students involved in the system change. The ACTIVITIES system is part of the TUTOR subsystem of the DANTE project - *Socio-Cultural Models implemented through multi-agent architecture for e-learning*. DANTE has as main objective the development of a model for the virtual education system, student centred, that facilitates the learning through collaboration as a form of social interaction.
Keywords: multi-agent system, e-learning, wasp models, tutor system, task allocation

1 Introduction

In [2] we have presented the DANTE *e-Learning* system. DANTE has an architecture with three levels (user, intermediary, supplier educational space), to each corresponding heterogeneous families of human agents and software. In the proposed system human agents interact with the artificial ones.

The *teacher* (human agent) is assisted by two types of software agents: *personal assistant* (classic interface agent) and *didactic assistant*. The student (human agent) evolves in an agentified environment with three types of agents. He/she also has a personal assistant (software interface agent) who monitors all the students' actions and communicates (interacts) with all the other agents, with the agentified environments of other students and the TEACHER agentified environment. The *student* has at his/her disposal two more agents: *TUTOR* and the *mediating agent*. The TUTOR assistant evaluates the educational objectives of the student and recommends her/him some kind of activities. The decisions are based on the knowledge of the students' cognitive and behavioural profiles (which takes into account the social component). The TUTOR multi-agent interacts with the personal assistant of the student, with the mediating agent and with the social agentified environment.

In this paper we are presenting a possible model for one of the components of the TUTOR sub-system, i.e. the one responsible for organizing the group-work activities for the students. A group-work activity can be, for example, a project that will be developed by a team of students. In a team, students are organized according to their skills and preferences. Students' skills are assessed by exams. For each exam the scores have a threshold and the student pass the exam only if the obtained score is greater than the threshold. The qualification of the student i for the course j is computed as follow:

$$q_{i,j} = \frac{p_{i,j} - c_{j,min}}{c_{j,max} - c_{j,min}} \tag{1}$$

where $p_{i,j}$ is the score obtained by the student i at the course j. $c_{j,min}$ is the minimum score to pass the exam and $c_{j,max}$ is the maximum score that can be obtained. In order to perform a specific activity the student needs to have the required qualifications. An activity consists of one or more tasks. The complexity of an activity is given in complexity points associated to component tasks. If an activity requires a qualification in more than one course, an average qualification score for all the required courses is used.

A student can be involved in many tasks of different activities, but the total number of complexity points for these tasks must not exceed a given maximum value. That means that in a student's activity queue we can have many tasks. The time period for every activity is strictly specified.

2 The ACTIVITIES System

In this paper we are presenting a multi-agent virtual environment - **ACTIVITIES**, that is a component of the TUTOR subsystem- where agents use wasp task allocation behaviour, combined with a model of wasp dominance hierarchy formation, to determine which task of a set of activities should be accepted into a student's queue, such

that the execution time of every activity will be respected and a maximum number of students will be involved in these activities. Wasp-like computational agents that were called learning routing wasps act as overall student proxies. The learning routing wasps must decide when to bid or when not to bid for arriving tasks. The environment is a dynamic one as new activities and new qualified students appear in time. The model presented here is similar with the one we have used for the assignment of activities in the GRANT sub-system of the **TUTOR** [1].

The wasp colony model has been successfully used in several allocation problems. The model of self-organization within a colony of wasps has been presented by Theraulaz et al.in [9]. In a colony of wasps, individual wasp interacts with its local environment in the form of a stimulus-response mechanism, which governs distributed task allocation. An individual wasp has a response threshold for each zone of the nest. Based on a wasp's threshold for a given zone and the amount of stimulus from brood located in this zone, a wasp may or may not become engaged in the task of foraging for this zone. A lowest response threshold for a given zone amounts to a higher likelihood of engaging in activity given a stimulus.

The goal of the **ACTIVITIE** System is to dynamically allocate the tasks of several activities from many activities, to qualified students, such that the time period allocated to every activity be respected and the number of students involved, maximized. Here is the model.

The student i has one or more course qualifications $q_{i,j}$ given by (1) and can be involved in various tasks from activities such that the sum of the complexity points for these activities must not exceed a limit value:

$$MCPS = Maximum\ Complexity\ Points/student \tag{2}$$

Each task of an activity has associated a number of complexity points and a set of courses which define the requirements for the task.

$$ncp_{j,k}\ - number\ of\ complexity\ points\ for\ the\ task\ j\ in\ the\ activity\ k \tag{3}$$

$$A_{i,j} = \{i_{1,j,k}, \ldots, i_{nj,k,j,k}\} \tag{4}$$

is the set of indexes of courses required by the task j from the activity k.

The minimum and maximum score that allow a student to perform the task j, from the activity k are:

$$min_{j,k} = \sum_{l \in A_{j,k}} c_{l,min}/N_{j,k} \tag{5}$$

where $N_{j,k}$ is the number of courses of $A_{j,k}$.

$$max_{j,k} = \sum_{l \in A_{j,k}} c_{l,max}/N_{j,k} \tag{6}$$

The tasks are classified by type. First, in the system are introduced a number of tasks' types, characterized by the sets T_m, which contain the courses that were required by these types.

$$T_m = \{c_{i1}, \ldots, c_{ikm}\} \tag{7}$$

The intersection of two sets of this kind has the following property:

$$\#(T_m \bigcap T_n) \leq p\%, min(\#T_m, \#T_n) \tag{8}$$

Each activity belongs to an unique type. The value p is a dynamical system parameter, that is modified in such a way that every task that is in the system in every moment, belongs to an unique task type set.

We will denote by

$$NT(t) = \#T = \#\{T_1, \ldots\}\ at\ the\ moment\ t \tag{9}$$

and by

$$ta_{j,k} = m \tag{10}$$

the type of the task j from the activity k (the task j, from the activity k is of type T_m). To each student i we associate two sets of indexes:

$$M_i = \bigcup_{student\ i} ta_{j,k} \tag{11}$$

is the set of all types of tasks in which she/he is or was involved.

$$M_{i,j} \subseteq M_i \tag{12}$$

is the set of all types of tasks she/he had already accomplished. The qualification of a student for the task j, from the activity k is

$$qa_{i,j,k} = average\{q_{i,j} | l \in A_{j,k}\} \tag{13}$$

If the incoming flow of new tasks allows, each of the students should specialize to one or more types of tasks among the ones she/he is capable to do. To model this requirement, we introduce the task specialization of a student. It takes into account the qualification of the student for the courses required by this task and the participation to other likewise tasks.

$$S_{i,j,k} = \omega \cdot qa(i,j,k) + \alpha \cdot \sum_{j \in M_i \cap \{ta_{j,k}\}} qa(i,j,k) + \beta \sum_{j \in M_{i,l} \cap \{ta_{j,k}\}} qa(i,j,k), \tag{14}$$

where ω, α, β are parameters of the system and have a major role in modelling of what "specialization" must represent. For the moment these parameters are tuned by hand. If we choose $\omega = 1,\ \beta > 1$ it means that the experience of student is more important than the initial score from different courses required by the tasks. If we choose $\alpha = 0,\ \beta = 0,\ \omega = 1$ it means that only the initial qualification is taking into account.

We denote by

$$sq_i - \ the\ number\ of\ tasks\ in\ the\ queue\ of\ the\ student\ i \tag{15}$$

This number satisfies the restriction:

$$\sum_{l=1}^{sq(i)} ncp_{l,k1} \leq MCPS \tag{16}$$

To each student we assigned a *learning routing* wasp that will be in charge to select which task to bid for entering the student's tasks' queue. In the same way as in the original wasp behaviour model, the learning routing wasp has a set of response thresholds associated to each tasks' type:

$$W_i = \{w_{i,j,k}\} \tag{17}$$

where $w_{i,j,k}$ is the response threshold of the wasp associated to the student i, for the task j from the activity k. The threshold value $w_{i,j,k}$ may vary in the interval $[w_{min}, w_{max}]$. Learning routing wasp will receive from tasks that have not been assigned to students a stimulus $S_{j,k}$ proportional to the waiting time for assignment. A learning routing wasp i will bid for a task j only if

$$\sum_{l \in A_{j,k}} q_{i,l}/N_{j,k} \geq min_{j,k} \tag{18}$$

The learning routing wasp will bid then for the task with a probability:

$$P(i,j,k) = \frac{S_{j,k}^{\gamma}}{S_{j,k}^{\gamma} + w_{i,j,k}^{\gamma}} \tag{19}$$

In (19) γ is a system parameter. Theraulaz et al. [9]have used a value of 2 for this parameter. In order to adjust the response thresholds for different tasks, each learning routing wasp, knows for every time moment the status of the student's queue, the characteristics of the task that is performed (the variables associated to activity, that is $A_{j,k}$, $ncp_{j,k}$, $qa_{i,j,k}$, $ta_{j,k}$, and if the student's status is idle). This update occurs at each time step.

3 Conclusions

In this paper we have presented a multi-agent virtual environment - ACTIVITIES, that is a component of a TUTOR subsystem- where agents use wasp task allocation behaviour, combined with a model of wasp dominance hierarchy formation, to determine which task of a set of activities should be accepted into a student's queue, such that the execution time of every activity will be respected and a maximum number of students will be involved in these activities. Wasp-like computational agents that were called learning routing wasps act as overall student proxies. The learning routing wasps must decide when to bid or when not to bid for arriving tasks. The environment is a dynamic one as new activities and new qualified students appears in time. The main characteristics of our system, which differentiates it from the system in [5], is that the number of tasks' type dynamically changes in time and that the number of tasks from a student queue depends on the restriction (16).

Further developments will aim to introduce decision mechanisms for establishing which learning routing wasp from a group of competing wasps gets a certain task, based on students' qualifications and to integrate this model with the student's model developed in the frame of the DANTE project [2].

Acknowledgement

This work benefits from founding from the research grants of the Romanian Ministry of Education, Research and Youth, code CNCSIS 33/2007 and CEEX 73/INFOSOC/2006.

References

[1] D. Simian, C. Simian, I. Moisil and I. Pah, Computer Mediated Communication and Collaboration in a Virtual Learning Environment Based on a Multi-agent System with Wasp-Like Behavior, in I. Lirkov, S. Margenov, and J. Wa´sniewski (Eds.): *LSSC 2007, LNCS* 4818, pp. 618-625, 2008, Springer-Verlag Berlin Heidelberg 2008

[2] I. Moisil, et al.: Socio-cultural modelling of student as the main actor of a virtual learning environment. *WSEAS Transaction on Information Science and Applications* 4, 30-36, 2007

[3] S. Goss, S. Aron, J. L. Deneubourg, J. M. Pasteels, Self-Organized Shortcuts in the Argentine Ant, *Naturwissenschaften*, 76:579-581, 1989.

[4] N. R. Jennings, M. Wooldridge, Applications of Intelligent Agents, in: N. R. Jennings and M. Wooldridge (editors), *Agent Technology: Foundations, Applications, and Markets*, Springer-Verlag, Heidelberg, Germany, 1998.

[5] V. A. Cicirelo, S.F. Smith,: Wasp-like Agents for Distributed Factory coordination. Autonomous Agents and Multi-Agent Systems 8(3), 237-267 (2004)

[6] E. Bonabeau, M. Dorigo, G. Theraulaz, Inspiration for optimization from social insect behaviour. *Nature*, vol. 406, 6 July 2000

[7] D. E. Goldberg, K. Deb, A comparative analysis of selection schemes used in genetic algorithms (1991),in FOGA91, vol. 1, pp 69-93.

[8] T. Stützle and M. Dorigo. ACO Algorithms for the Travelling Salesman Problem. In *Proceedings of the EUROGEN conference*, M Makela, K Miettinen, P Neittaanmaki, J Periaux (Eds), John Wiley and Sons, ISBN: 0471999024, 1999.

[9] G. Theraulaz, et al., Task differention in policies waspcolonies. A model for selforganizing groups of robots. In: *From animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive behavior*, pp. 346-355, 1991.

Ioana Moisil, Dana Simian
"Lucian Blaga" University of Sibiu, Romania
4, Emil Cioran Str., Sibiu 550025, Romania
E-mail: ioana.moisil@ulbsibiu.ro

Iulian Pah
"Babes-Bolyai" University Cluj-Napoca, Romania

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 417-421

Piezo Smart Wing with Sliding Mode Control

Eliza Munteanu, Ioan Ursu, Aurel Alecu

Abstract: The objective of the present work is the investigating of the capabilities of piezoelectric actuators as active vibration control devices for wing structures. Herein, the enhancing of the wing dynamic behavior is then based on the sliding mode control synthesis, having in view its opportunities in the performing of robust systems. Numerical simulations are presented, showing the efficacy of the piezo actuators in the developing of smart structures.
Keywords: smart structure, wing, sliding mode control, numerical simulation.

1 Introduction

Certification regulations require that any certified aircraft is free of wing dangerous vibrations. The active techniques enhance dynamic behavior of the wing, without redesign and adding mass; nowadays, these are used both for flutter suppression and structural load alleviation. Thus, our target concerns the sliding mode control synthesis, for a piezo smart composite wing. The approach continues recent researches of the authors, see [1], [2].

2 The Mathematical Model

The computational program ANSYS, performing FEM analysis of a wing physical model (Figure 1) defined only in terms of geometrical and structural data, was applied to obtain the structural, order two, mathematical model

$$M\ddot{x}+C\dot{x}+Kx=0, \tag{1}$$

where x is the vector of nodal displacements, and M, C and K are mass, damper and stiffness matrices. The wing skin is built on composite material E-glass texture/orto-ophthalic resin with 3 layers and 0.14 mm thickness each of them. The wing spars, placed at 25%, respectively, 65% of chord, are performed of dural D16AT (1 layer, with 5 mm thickness). The interior of the wing is filled with a polyurethane foam. The ANSYS geometric model equipped with MFC actuators is given in Figure 1.

Following a modal analysis using the full ANSYS model, the first four natural modes and frequencies (Hz) — 8.33; 20.55; 94.38; 131.97 — were found. Then, by an ANSYS substructuring analysis, the mathematical model was completed in the form

$$M\ddot{x}+C\dot{x}+Kx=\tilde{B}_1\xi+\tilde{B}_2u, \tag{2}$$

where $\tilde{B}_1$, $\tilde{B}_2$ are the matrices of the influences of the perturbation ξ and the control u. The operation assumed the static interaction cause-effect

$$Kx_k=\tilde{B}_2u_k, \qquad Kx_k=\tilde{B}_1\xi_k, \tag{3}$$

where x_k is the displacement vector corresponding to a unitary electric field u_k applied to the k MFC actuator, herein $k=1,2$; the two columns of the matrix $\tilde{B}_2$ are so obtained. To calculate piezo action, we used the analogy between thermal and piezoelectric equations developed in [3]. Analogously one proceeds for obtaining of the matrix $\tilde{B}_1$, by applying the unitary force ξ_k in the point 3 noted in Fig. 1. The subsequent operations concern a) the recuperation in MATLAB of the matrices in system (2) and b) the modal transforming

$$x=Vq \tag{4}$$

of the system (2) by using a reduced modal matrix (of order four) of eigenvectors V of dimension 34281×4 (34281 is the number of generalized coordinates in ANSYS, in connection with the number of the chosen FEM nodes)

$$V^{\mathrm{T}}MV\ddot{q}+V^{\mathrm{T}}KVq=V^{\mathrm{T}}\tilde{B}_1\xi+V^{\mathrm{T}}\tilde{B}_2u. \tag{5}$$

Thus, modal quasidecentralized system, of four modes, is obtained (a distribution of critical dampings ζ_i is chosen)

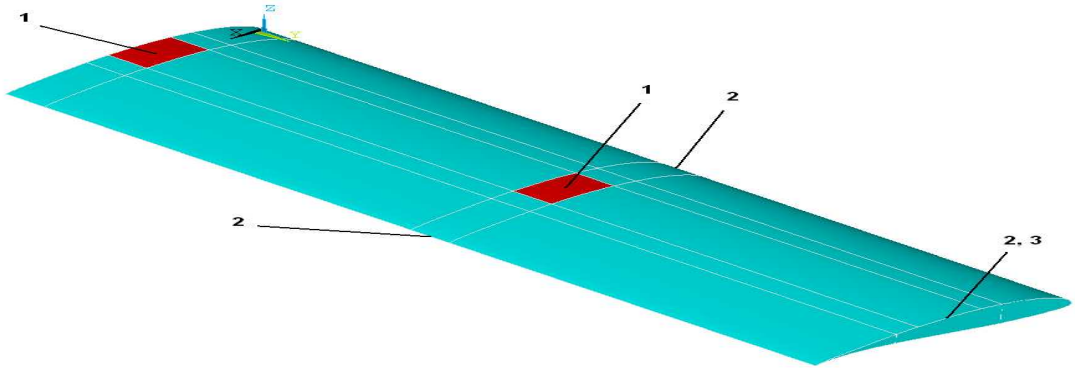


Figure 1: ANSYS model for wing equipped with MFC actuators (1) and sensors (2). A simple perturbation, as representing the aerodynamic forces, is considered, see (3).

$$\ddot{q}+\text{diag}\,(2\zeta_i\omega_i)\dot{q}+\text{diag}\left(\omega_i^2\right)q=B_1\xi+B_2u,\ \zeta_i\,[I_1\ I_2\ I_3\ I_4]:=[0.04\ 0.03\ 0.025\ 0.02]\,, \tag{6}$$

for $i=1,\ldots,4$, as a basis for defining the first order state system

$$\dot{x}(t)=Ax(t)+B_1\xi(t)+B_2(t)u(t),\quad z(t)=C_1x(t),\quad y(t)=C_2x(t)+\mu I\eta(t), \tag{7}$$

where $x(t)$ is the state, $z(t)$ is the controlled output, $y(t)$ is the measured output, and $u(t)$ is the control input. The state vector is given by

$$x(t)=(q_4,\ q_3,\ q_2,\ q_1,\ \dot{q}_4,\ \dot{q}_3,\ \dot{q}_2,\ \dot{q}_1)^{\mathrm{T}}. \tag{8}$$

The external and internal components of the perturbations ξ and η are the substitute of the aerodynamic disturbances and sensor noise vector, respectively. The controlled output $z(t)$ will concern the whole system state see matrix C_1 below. As the measured output $y(t)$ is taken the z-axis components of the generalized coordinates associated to nodes noted in Figure 1. Consequently, the system's matrices is succinctly transcribed

$$C_1=\left[\text{diag}\left(1,10^2,10^4,1\right)\ \text{diag}\left(10,10^4,10^5,10^7\right)\right],\ C_{2,3\times8}:=C_{2,3\times34281}V_{34281\times4}\,[I_4\ 0_{4\times4}]\,. \tag{9}$$

3 The Framework of LQG Control Synthesis

The LQG control synthesis concerns the system (7). The goal is to find a control such that the system is stabilized and the control minimizes the cost function, relatively to $z(t)$ and $u(t)$,

$$J_{LQG}=\lim_{T\to\infty}\mathrm{E}\left\{\int_0^T\begin{bmatrix}x(t)^{\mathrm{T}} & u(t)^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}C_1^{\mathrm{T}}C_1 & 0\\ 0 & \rho_RI_2\end{bmatrix}\begin{bmatrix}x(t)\\ u(t)\end{bmatrix}dt\right\}. \tag{10}$$

The solution consists in the building of a controller and a state-estimator (Kalman filter) [4].

Now, a LQG/LTR procedure is proposed in view of obtaining a comparison term for the sliding mode control presented in Section 4. The H_2-tradeoff type optimization perspective used in [5] is considered. Thus, the controller gain synthesis is performed such that

$$C_1\,(sI-A)^{-1}B_2/\rho_R=W(s), \tag{11}$$

where $W(s)$ is a weight chosen such that its crossover frequency be at least the frequency of the first neglected mode (the five one, herein). The filter gain synthesis will be performed such that

$$B_1\equiv B_2,\qquad \mu\to0,\qquad L_{\mathrm{LQG}}(j\omega)\approx L_{\mathrm{LQR}}(j\omega) \tag{12}$$

in a certain range $\omega \in [0, \omega_{\max}]$, as large as possible, where $(j\omega = s)$

$$\begin{aligned} L_{\mathrm{LQG}}(s) &= \left[-\left(sI - A - K_f C_2 - B_2 K_c\right)^{-1} K_f C_2 \right] (sI - A)^{-1} B_2, \\ L_{\mathrm{LQR}}(s) &= K_c \, (sI - A)^{-1} B_2. \end{aligned} \tag{13}$$

Herein, the filter gain K_f was chosen so that the closed-loop LQG/LTR system (having open loop matrix L_{LQG}) recovers internal stability and some of the robustness properties (gain and phase margins) of the LQR design (with open loop matrix L_{LQR}).

4 Sliding Mode Control Synthesis

The sliding mode [6]–[10] with m manifolds (m is the dimension of the control vector u)

$$H_i := g_i^{\mathrm{T}} \hat{x} = 0, \qquad i = 1, 2, \ldots, m, \tag{14}$$

is thought so that the system be stable as long as the state remains on the hyperplane (14). The equivalent control law to keep the state on the hyperplane is given by

$$u_{eq} = -\left(GB_2\right)^{-1} \left[G\left(A - K_f C_2\right) \hat{x} + G K_f y \right], \quad G := [g_1, g_2, \ldots, g_m]^{\mathrm{T}} \tag{15}$$

which means: if the estimated states never leave sliding hyperplane H_i, then $dH_i/dt = 0$, for $i = 1, \ldots, m$. To satisfy the reaching condition (the condition for any initial state to reach the sliding manifold) the control law is chosen as

$$u = u_{eq} - \left(GB_2\right)^{-1} \operatorname{diag} \beta \operatorname{sgn} [H_1, H_2, \ldots, H_m] \tag{16}$$

which means: the Lyapunov function $V = H^2/2$, $H := [H_1, H_2, \ldots, H_m]$, has negative derivative, $dV/dt < 0$, and the state of the system is transferred on the hyperplane H_i (sgn is a notation for the signum function). $\operatorname{diag}(\beta)$ is a diagonal matrix with the ith diagonal entry equal to a positive number β_i, $i = 1, \ldots, m$. To eliminate the chattering behaviour, the saturation function will substitute the signum function in (16)

$$\operatorname{sat} H_i = \begin{cases} \operatorname{sgn} H_i & \text{if} \quad |H_i| > \delta_i \\ H_i/\delta_i & \text{otherwise}, \quad i = 1, \ldots, m. \end{cases} \tag{17}$$

Thus, the effective linear control law can finally be written as (n is the dimension of the state)

$$u = -\left(GB_2\right)^{-1} \left[G\left(A - K_f C_2 + \rho I_n\right) \hat{x} + G K_f y \right], \tag{18}$$

where, without loss of generality, β_i and δ_i have been assumed to be β and δ and $\rho := \beta/\delta$. The vectors g_i will be sought to minimise a reduced quadratic objective function versus (10),

$$J := \int_0^\infty x^{\mathrm{T}} Q x dt, \qquad Q = C_1^{\mathrm{T}} C_1 \geq 0. \tag{19}$$

Let the matrix P be composed of basis vectors of the null space of B_2^{T}, $\ker(B_2^{\mathrm{T}})$. Defining a similarity transformation

$$q(t) = Gx(t), \quad G := [P \; B_2]^{\mathrm{T}} \tag{20}$$

and ignoring the disturbance ξ, the first equation in (7) can be recast as

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} u. \tag{21}$$

Thus

$$J = \int_0^\infty q^{\mathrm{T}} \left\{ \left(G^{-1}\right) \right\}^{\mathrm{T}} Q G^{-1} q dt := \int_0^\infty \left(q_1^{\mathrm{T}} Q_1 q_1 + q_1^{\mathrm{T}} N q_2 + 2 r q^2 \right) dt. \tag{22}$$

The equation

$$\dot{q}_1 = S_{11} q_1 + S_{12} q_2 \tag{23}$$

and (22) represent a standard linear quadratic problem provided $r > 0$ (if not, a new Q will be chosen, see (9). Now, if H is expressed as

$$H = Kq_1 + q_2 \tag{24}$$

then $H = 0$ implies $q_2 = -Kq_1$ and in view of equation (23), K is a state feedback gain vector. For a full-state regulator problem, the equation (14) will take the form of $H = GX$; thus

$$\begin{gathered} G = [K \quad I_n]H, \qquad K = R^{-1}\left[S_{12}^{\mathrm{T}}P_2 + N^{\mathrm{T}}\right], \\ P_2\left(S_{11} - S_{12}R^{-1}N^{\mathrm{T}}\right) + \left(S_{11}^{\mathrm{T}} - NR^{-1}S_{12}^{\mathrm{T}}\right)P_2 - P_2S_{12}R^{-1}S_{12}^{\mathrm{T}}P_2 + Q_1 - NR^{-1}N^{\mathrm{T}} = O. \end{gathered} \tag{25}$$

5 Numerical Application and Conclusions

The afore described sliding mode control was proven in numerical simulations using as comparison terms the passive system

$$\dot{x}(t) = Ax(t) + B_1\xi(t) \tag{26}$$

and the LQG/LTR system defined in Section 3. Both the LQG/LTR and sliding mode systems were taken into account with comparable energy of control vector and have the same filter gain.

By *a trail and error process*, the suitable values of LQG/LTR, respectively, sliding mode weighting matrices were chosen

$$Q_\xi = q_\xi I_2,\ Q_\eta = \mu^2 I_3,\ q_\xi = 5\times 10^5,\ \mu = 10^{-8},\ \rho_R = 5\,000,\ \rho = 1. \tag{27}$$

The state perturbation

$$\xi_1 = 10^8\cos(2\pi\times 20.55t), \xi_2 = 10^6\cos(2\pi\times 94.38t) + 10^6\cos(2\pi\times 131.97t) \tag{28}$$

was considered as generating a system resonance for the most dangerous mode two. Similar sensor noises were introduced

$$\eta_1 = 0.1\cos(2\pi\times 50t), \quad \eta_2 = 0.1\cos(2\pi\times 100t), \quad \eta_3 = 0.1\cos(2\pi\times 175t). \tag{29}$$

With the relations

$$\delta(q_{i,LTR}) = \frac{std(q_{i,P}) - std(q_{i,LTR})}{std(q_{i,P})}\times 100, \delta(q_{i,SM}) = \frac{std(q_{i,P}) - std(q_{i,SM})}{std(q_{i,P})}\times 100, \tag{30}$$

for $i = 1,\ldots,4$, were calculated percentual deviations of the LQG/LTR ("LQG") modes rms values versus the passive ("P") modes rms values, and similar percentual deviations of the sliding mode ("SM") modes rms values. The numerical results for $\delta(q_{i,LTR})$, were: 29.22, 26.02, 23.15, 17.99 percents, respectively 92.00, 98.16, 17.20, 29.02 percents for sliding mode case. One can see a better vibration attenuation in the case of sliding mode controller versus the LQG/LTR controller for the nominal case, excepting the attenuation of the mode 3. Worthy mentioning, the vibration control focused on the attenuation of the most dangerous first bending mode, the mode 2. The nominal controllers were further considered in a comparative robustness evaluation. The Table 1 summarizes the main conclusions concerning the robustness properties of the two controllers in the case of natural frequencies uncertainties. Percentual modes vibration attenuation performances were calculated by the relation

$$\frac{std(q_{i,LTR}) - std(q_{i,SM})}{std(q_{i,LTR})}\times 100. \tag{31}$$

A first conclusion concerns a favorable influence on the sliding mode controller robustness in the case of a conjugated uncertainty increasing in positive values on the modes 1 and 2 and in negative values on the modes 3 and 4 (see the 5^{th} row in the Table 1). A second conclusion concerns a relatively favorable influence on the controller robustness of a conjugated increasing in negative values of the modes 1 and 4 uncertainty (see the 6^{th} row in the Table 1). The 7^{th} row show unfavorable configurations of frequency uncertainties. It is the single considered case when the sliding mode controller works worse than the open loop system.

Thus, the main conclusion of the work concern the idea of validating the proposed LQG/LTR sliding mode control as an effective methodology of vibrations attenuation for a piezo smart wing.

Table 1: Influence on the robustness of the frequency uncertainty

percents deviations for mode I, [%] see the relation (31)				random uncertainties on mode i natural frequency — deviation versus nominal values, [%]			
$i=1$	$i=2$	$i=3$	$i=4$	$i=1$	$i=2$	$i=3$	$i=4$
86.46	86.90	2.79	9.94	17	–15	15	–17
83.58	83.35	**–0.43**	5.78	20	–20	15	–15
78.78	77.34	**–11.13**	**–1.46**	25	–25	20	–20
60.61	92.63	33.77	**–8.53**	20	–10	–20	10
84.74	85.01	45.17	27.63	20	25	–20	–25
82.32	85.34	**–25.62**	21.64	–20	20	25	–25
-10^8	-10^8	-10^8	-10^8	–10	20	–20	10

References

[1] L. Iorga, E. Munteanu, I. Ursu, "Enhancing wing dynamic behavior by using piezo patches," *Proceedings of International Conference in Aerospace Actuation Systems and Components*, Toulouse, June 13-15, pp. 171–176, 2007.

[2] E. Munteanu, I. Ursu, "Robust LQG/LTR control synthesis for flutter alleviation," *ICTAMI 2007, The International Conference on Theory and Applications of Mathematics and Informatics*, 29 August-2 September 2007, Alba Iulia, Romania.

[3] N. Mechbal, "Simulations and experiments on active vibration control of a composite beam with integrated piezoceramics," *Proceedings of 17th IMACS World Congress*, 2005, France.

[4] R. E. Kalman, "Contributions to the theory of optimal control," *Bol. Soc. Mat. Mexicana*, **5**, pp. 102–109, 1960.

[5] G. Stein, M. Athans, "The LQG/LTR procedure for multivariable feedback control design," *IEEE Trans. Automat. Control*, **32**, pp. 105–114, 1987.

[6] V. I. Utkin, "Variable structure system with sliding mode: a survey," *IEEE Trans. Automat. Control*, **22**, pp. 212–222, 1977.

[7] A. Sinha, D. W. Miller, "Optimal sliding-mode control of a flexible spacecraft under stochastic disturbance," *J. Guidance, Control Dyn.*, **18**, pp. 486–492, 1995.

[8] Ursu, F., T. Sireteanu, I. Ursu, "On anti-chattering synthesis for active and semi-active suspension systems," *Preprints of the* 3^{rd} *IFAC International Workshop on Motion Control*, Grenoble, France, September 21-23, pp. 93–98, 1998.

[9] J.-J. Slotine, "Sliding controller design for nonlinear systems," *Int. J. Control*, **40**, pp. 421–434, 1984.

[10] I. Ursu, F. Ursu, "Active and semiactive control," *Romanian Academy Publishing House*, 2002.

Eliza Munteanu
Advanced Studies and Research Center
Bucharest, Romania
E-mail: elizamun@gmail.com

Ioan Ursu
"Elie Carafoli" National Institute for Aerospace Research
Bucharest, Romania
E-mail: iursu@aero.incas.ro

Aurel Alecu
University Politehnica of Bucharest, Department of Mechanics
Bucharest, Romania
E-mail: alecu@cat.mec.pub.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 422-426

Generic Procedure for Construction of a Multi-Dimensional Utility Function under Fuzzy Rationality

Natalia Nikolova, Sevda Ahmed, Kiril Tenekedjiev

Abstract:
The paper presents a generic procedure for construction of multi-dimensional utility in the case of mutual utility independence of base vector attributes, and of fuzzy rationality (i.e. when preferences of the decision maker only partially obey the rationality axioms). The procedure incorporates techniques for direct elicitation of multi-dimensional utilities, techniques for elicitation of one-dimensional utility, procedures for utility analysis of non-monotonic preferences, and the uniform method for finding point estimates of interval scaling constants of the multi-dimensional utility function.
Keywords: utility function, fuzzy rationality, multi-dimensional prizes, attributes

1 Introduction

Utility theory [von Neumann, Morgenstern, 1947] applies in choices under risk between alternatives (lotteries), each generating one out of a set X of consequences (prizes) with a known probability. The utility function $u(.)$ is the measure of preferences of the decision maker (DM) over the prizes. The construction of $u(.)$ is based on rationality axioms [French, Insua, 2000], which assume that utility might be elicited by solving a preferential equation of lotteries and/or prizes via a dialog with the DM. The ideal DM obeys the rationality axioms and elicits point estimates of utilities. The real DM has finite discriminating abilities and elicits uncertainty intervals. As a result she partially disobeys some of the rationality assumptions, and is referred to as fuzzy-rational (FRDM) [Nikolova, et al., 2005].

In some cases X is a one-dimensional set of prizes and preferences are monotonic. Then it is recommended to elicit several points of the curve and approximate/interpolate the utility. Classical and modern elicitation techniques are probability equivalence (PE) [Clemen, 1996], certainty equivalence (CE) [Farquhar, 1984], lottery equivalence (LE) [McCord, De Neufville, 1986], trade-off (TO) [Wakker, Deneffe, 1996], and uncertain equivalence (UE) [Tenekedjiev, et al., 2006] methods. Further approximation of utility requires a choice of an analytical form depending on the risk attitude of the DM and the type of prizes [Keeney, Raiffa, 1993]. In the case of non-monotonic preferences, the elicitation techniques can only be applied if the extrema are identified [Nikolova, et al., 2006].

The objectives that a DM has in a given situation lead to identification of multi-dimensional prizes, whose coordinates measure the degree of achievement of each objective. There is a possibility to analyze such prizes as a whole, but results are likely to be biased. Therefore it is recommended to decompose the multi-dimensional $u(.)$ to fundamental utility functions over each attribute (or subgroups of attributes). Independence conditions ensure that such a representation corresponds to the opinion of the DM. It is recommended to define mutual utility independence and represent the multi-dimensional $u(.)$ as a combination of the fundamental utility functions and their scaling constants. The sum of these constants (either 1 or not) defines whether an additive or a multiplicative multi-dimensional $u(.)$ shall be constructed. The scaling constants are subjectively elicited, and thus also affected by fuzzy rationality [Nikolova, 2007a].

The paper summarizes the theoretical aspects that comprise the construction of a multi-dimensional utility function of a FRDM. Since decision situations differ in structure and complexity, the review ends up with the elaboration of a 6-step algorithm. It describes the stages of construction of a multi-dimensional utility under mutual utility independence taking into account the complexity and dimensionality of prizes, the types of preferences, and the fuzzy rationality of the DM. In what follows, section 2 is a review of techniques for construction of one-dimensional $u(.)$ by a FRDM, whereas section 3 does the same for the case of multi-dimensional $u(.)$. All these techniques combine to form the 6-step algorithm for construction of multi-dimensional utilities under fuzzy rationality in section 4.

2 Utility function in the one-dimensional case

A classical task in decision analysis is to build a utility function over a one-dimensional set X. The function is constructed in the interval $[x_{worst}; x_{best}]$, where $x_{best} = sup(X)$, $x_{worst} = inf(X)$. The procedures depend on the

type of preferences of the DM.

In most case, preferences are monotonically increasing, i.e. $x_i \succ x_j \Leftrightarrow x_i > x_j, x_i \in X, x_j \in X$. It is reasonable to elicit only several (z) knots and construct $u(.)$ using approximation/interpolation. Depending on the elicitation technique, the resulting estimates will be uncertainty intervals either on the prizes or on the utilities. Once elicited knots are available, the utility function may be constructed using linear interpolation or analytical approximation. Analytical approximation of an monotonically increasing one-dimensional $u(.)$ is discussed in [Tenekedjiev, et al., 2007]. The selected method should precisely interpret the data, and should preserve the risk attitude of the DM [French, 1993], modeled by the local risk aversion function $r(x) = --u''(x)/u'(x)$ [Pratt, 1964]. A rich source of analytical forms is [Keeney, Raiffa, 1993]. The work [Stoianov, 1993] proposed the Harrington's desirability function, whereas [Nikolova, 2007b; Tenekedjiev, et al., 2007] discuss an arctg-approximation that applies over gains and loses, and encapsulates prior information for the risk attitude.

A more complex situation arises when the DM has quasi-unimodal preferences, i.e. when there is an extreme point x_{opt} within the interval of prizes [Nikolova, et al., 2006]. There are two types of quasi-unimodal preferences - hill (with a maximum extremum) and valley (with a minimum extremum) preferences. Both occur due to two contradicting factors related to the analyzed variable. Difficulties arise when the DM has to compare values on both sides of the extreme interval. This leads to mutual non-transitivity of preferences and to fuzzy rationality. As a result, she would identify an extreme interval of x_{opt}. The models of hill and valley utility functions are based on two separate sets of assumptions. Two 20-step algorithms are elaborated to find the extreme interval via a dialog with the FRDM. Both algorithms combine the golden section search [Kiefer, 1953] and bisection [Press, et al., 1992]. Golden section serves to locate the extreme interval, whereas bisection estimates its lower and upper bounds. After the extreme interval is identified, local utility functions are constructed in the sections with monotonic preferences (the sections of each side of the extremum) and then rescaled into the global utility. Two other algorithms are proposed for that purpose in [Nikolova, et al., 2006].

3 Utility function in the multi-dimensional case

DMs have many different objectives in a decision situation, which results in the construction of multi-dimensional consequences, represented as d-dimensional vectors, whose coordinates (attributes) measure the important aspects in the problem for the DM. If the i-th attribute is a random variable X_i with an arbitrary realization x_i, then prizes take the form of d-dimensional vectors $\overrightarrow{\mathbf{x}} = (x_1, x_2, \ldots, x_d)$, $\overrightarrow{\mathbf{x}} \in X^d$. The set $\{X_1, X_2, \ldots, X_d\}$ may be divided into $n \in \{2, 3, \ldots, d\}$ non-empty non-overlapping subsets $Y_1, Y_2, \ldots, Y_n$, called base vector attributes. Each Y_i is a system of random variables with an arbitrary realization $\overrightarrow{\mathbf{y}}_\mathbf{i}$ and then prizes in X may be presented as $\overrightarrow{\mathbf{x}} = (\overrightarrow{\mathbf{y}}_\mathbf{1}, \overrightarrow{\mathbf{y}}_\mathbf{2}, \ldots, \overrightarrow{\mathbf{y}}_\mathbf{n})$ [Nikolova, 2007a]. Theoretically speaking, it is possible to construct the multi-dimensional utility function $u(.)$, whose domain are all possible $\overrightarrow{\mathbf{x}} \in X^d$:

$$u = u(\overrightarrow{\mathbf{x}}) = u(x_1, x_2, \ldots, x_d). \tag{1}$$

An algorithm for that purpose was proposed in [Keeney, Raiffa, 1993], which, as the authors argue, becomes practically inapplicable if $d > 3$. A much better approach is to decompose (1) to base utility functions over $Y_1, Y_2, \ldots, Y_n$:

$$u(\overrightarrow{\mathbf{x}}) = u(\overrightarrow{\mathbf{y}}_1, \overrightarrow{\mathbf{y}}_2, \ldots, \overrightarrow{\mathbf{y}}_n) = f[u_1(\overrightarrow{\mathbf{y}}_1), u_2(\overrightarrow{\mathbf{y}}_2), \ldots, u_n(\overrightarrow{\mathbf{y}}_n)]. \tag{2}$$

Here, $f(.)$ is a real-valued function of n real variables. The adequacy of (4) requires that certain independence conditions of preferences over attributes hold. Preferential independence is most common and the weakest independence condition. The strongest condition is additive independence, but is very difficult to establish.

Of greatest importance is utility independence. Let's divide the set $\{Y_1, Y_2, \ldots, Y_n\}$ into two non-empty non-overlapping subsets Z and $\overline{Z}$, where $\overline{Z}$ is complementary to Z. Since Z and $\overline{Z}$ are random vectors, then these along with their realizations may be analyzed as vector attributes of a given prize: $\overrightarrow{\mathbf{x}} = (\overrightarrow{\mathbf{z}}, \overrightarrow{\overline{\mathbf{z}}})$. Let $l_{i, \overrightarrow{\overline{\mathbf{z}}}}$ be a specific lottery with prizes, for which $\overline{Z} = \overrightarrow{\overline{\mathbf{z}}}$. The vector attribute Z is utility independent from $\overline{Z}$, when the preference order over lotteries with fixed $\overline{Z}$ and variable Z does not depend on $\overline{Z}$, i.e. from $l_{i, \overrightarrow{\overline{\mathbf{z}}}_\mathbf{k}} \gtrsim l_{j, \overrightarrow{\overline{\mathbf{z}}}_\mathbf{k}} \Rightarrow l_{i, \overrightarrow{\overline{\mathbf{z}}}} \gtrsim l_{j, \overrightarrow{\overline{\mathbf{z}}}}, \forall \overrightarrow{\overline{\mathbf{z}}} \in \overline{Z}$. If that holds, then Z is utility independent. If each Z is utility independent, then $Y_1, Y_2, \ldots, Y_n$ are mutually utility independent. The most preferred value of the base vector attribute Y_i is $(\overrightarrow{\mathbf{y}}_i)_{best}$, whereas the least preferred is $(\overrightarrow{\mathbf{y}}_i)_{worst}$. Mutual utility independence allows to construct $u(.)$ over n arguments using n normalized base utility functions $u_i(.)(i = 1, 2, \ldots, n)$ [Keeney, Raiffa, 1993]:

$$u(\overrightarrow{\mathbf{x}}) = u(\overrightarrow{\mathbf{y}}_1, \overrightarrow{\mathbf{y}}_2, \ldots, \overrightarrow{\mathbf{y}}_n) =$$
$$= \sum_{i=1}^{n} k_i u_i(\overrightarrow{\mathbf{y}}_i) + k \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_i k_j u_i(\overrightarrow{\mathbf{y}}_i) u_j(\overrightarrow{\mathbf{y}}_j) +$$
$$+ k^2 \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{r=j+1}^{n} k_i k_j k_r u_i(\overrightarrow{\mathbf{y}}_i) u_j(\overrightarrow{\mathbf{y}}_j) u_r(\overrightarrow{\mathbf{y}}_r) + \cdots +$$
$$+ k^{n-1} k_1 k_2 \ldots k_n u_1(\overrightarrow{\mathbf{y}}_1) u_2(\overrightarrow{\mathbf{y}}_2) \ldots u_n(\overrightarrow{\mathbf{y}}_n), \quad (3)$$

In (3), $u(\overrightarrow{\mathbf{x}}_{\mathbf{best}}) = u[(\overrightarrow{\mathbf{y}}_1)_{best}, (\overrightarrow{\mathbf{y}}_2)_{best}, \ldots, (\overrightarrow{\mathbf{y}}_n)_{best})] = 1$, $u(\overrightarrow{\mathbf{x}}_{\mathbf{worst}}) = u[(\overrightarrow{\mathbf{y}}_1)_{worst}, (\overrightarrow{\mathbf{y}}_2)_{worst}, \ldots, (\overrightarrow{\mathbf{y}}_n)_{worst})] = 0$, $u_i[(\overrightarrow{\mathbf{y}}_1)_{best}] = 1$ $(i = 1,2,\ldots,n)$, and $u_i[(\overrightarrow{\mathbf{y}}_1)_{worst}] = 0$ $(i = 1,2,\ldots,n)$. The constants in (3) are called scaling constants, and if $\overrightarrow{\mathbf{x}}_{\mathbf{i,corner}} = [(\overrightarrow{\mathbf{y}}_1)_{worst}, (\overrightarrow{\mathbf{y}}_2)_{worst}, \ldots, (\overrightarrow{\mathbf{y}}_i)_{best}, \ldots, (\overrightarrow{\mathbf{y}}_n)_{worst}]$ is a corner consequence, then $k_i = u[(\overrightarrow{\mathbf{y}}_1)_{worst}, (\overrightarrow{\mathbf{y}}_2)_{worst}, \ldots, (\overrightarrow{\mathbf{y}}_i)_{best}, \ldots, (\overrightarrow{\mathbf{y}}_n)_{worst}]$, $1 + k = \prod_{i=1}^{n}(k \times k_i + 1)$.

The paper [Nikolova, 2007a] suggests a LE-like scheme of elicitation of scaling constants. As the DM is only fuzzy rational, her estimate takes the form $k_i \in [\hat{k}_{d,i}; \hat{k}_{u,i}]$ $(i = 1,2,\ldots,n)$, elicited using triple bisection [Tenekedjiev, et al., 2004]. The construction of $u(.)$ over the multi-dimensional consequences with n number of base vector attributes requires to find whether $k_1 + k_2 + \cdots + k_n = 1$, and then find point estimates of the constants $\hat{k}_i$, for $i = 1,2,\ldots,n$.

The uniform method has been proposed to solve the scaling constants' problem [Nikolova, 2007a; Tenekedjiev, 2008]. Assume that $a_n = \sum_{i=1}^{n} k_{d,i}$, $b_n = \sum_{i=1}^{n} k_{u,i}$, $y_n = \sum_{i=1}^{n} k_i$ $(i = 1,2,\ldots,n)$, $s = \frac{a_n + b_n}{2}$, and that each unknown constant is uniformly distributed in its uncertainty interval. If $m > 0$, $a_n < 1$, $b_n > 1$, then it is necessary to find the distribution law of y_n. An analytical procedure to construct the density $f_{y_n}(.)$ of the sum of scaling constants is proposed in [Tenekedjiev, 2008], as well as its numerical approximation. A Monte-Carlo based simulation approximation is presented in [Nikolova, 2007a], which finds the distribution law in the form of a cumulative distribution function $F_{y_n}^{n}(.)$.

The sum of the scaling constants cannot be deducted if $a_n < 1$, $b_n > 1$. For that case, Nikolova (2007a) uses a two-tail statistical test, where H_0 is $y_n = 1$, and H_1 is $y_n \neq 1$. At a level of significance α and calculated probability p_{value} to reject H_0 that is true, H_0 is rejected and H_1 is accepted if $p_{value} \leq \alpha$, or H_0 fails to be rejected if $p_{value} > \alpha$ [Hanke, Reitsch, 1991]. It is necessary to assess p_{value}. An appendix in Nikolova (2007) proves that $\hat{p}_{value} = 2\int_{a_n}^{1} \hat{f}_{y_n}(y)\mathrm{d}y$ if $s > 1$, and $\hat{p}_{value} = 2 - 2\int_{a_n}^{1} \hat{f}_{y_n}(y)\mathrm{d}y$ if $s \leq 1$ (s is the sum of the lower and upper bounds of the constants, divided by 2). These dependencies are numerically approximated in [Tenekedjiev, 2008], whereas a simulation approximation in [Nikolova, 2007a] proves that $\hat{p}_{value} = 2\hat{F}_{y_n}^{n}(1)$ if $s > 0.5$, and $\hat{p}_{value} = 2 - 2\hat{F}_{y_n}^{n}(1)$ if $s \leq 0.5$. Point estimates of the constants are defined depending on the result of the test, following (4), where $\beta = (b_n - -1)/(b_n - -a_n)$:

$$\hat{k}_i = \begin{cases} \beta k_{d,i} + (1-\beta) k_{u,i}, & \text{if } H_0 \text{ fails to be rejected} \\ (k_{d,i} + k_{u,i})/2, & \text{if } H_1 \text{ is accepted} \end{cases} \quad (4)$$

The sum of k_i defines the form of the utility function, as follows:

$$u(\overrightarrow{\mathbf{x}}) = \begin{cases} \sum_{i=1}^{n} k_i u_i(\overrightarrow{\mathbf{y}}_\mathbf{i}), & \text{if } \sum_{i=1}^{n} k_i = 1 \\ \dfrac{\prod_{i=1}^{n} \left[K k_i u_i(\overrightarrow{\mathbf{y}}_\mathbf{i}) + 1\right] - 1}{K}, & \text{if } \sum_{i=1}^{n} k_i \neq 1 \end{cases} \quad (5)$$

In (5), K is a general constant, which depends on the value of the scaling constants. An algorithm to find K is proposed by Keeney and Raiffa (1993).

4 Generic procedure for construction of multi-dimensional utility

All procedures that were discussed earlier may be united into a generalized algorithm to construct a multi-dimensional utility function of a FRDM.

Algorithm 1. *Constructing multi-dimensional utility function of a FRDM*

1. Establish mutual utility independence between the base vector attributes $Y_1, Y_2, \ldots, Y_n$.
2. Decompose each function $u_{y,j}(.)$ for $d_j > 1$ in case mutual utility independence was established for some of the attributes the function is defined on [Engel, Wellman, 2006].
3. Construct all non-decomposed $u_{y,j}(.)$ for $d_j > 1$ and all multi-dimensional decomposed parts of $u_{y,j}(.)$ using the algorithm from [Keeney, Raiffa, 1993].
4. Construct all one-dimensional utilities and one-dimensional decomposed parts of $u_{y,j}(.)$:
   (a) in the case of non-monotonic preferences: define the number of local extrema and divide the one-dimensional set of prizes into sectors with pseudo-unimodal preferences; elicit the uncertainty interval of the extremum in each sector using the procedures from [Nikolova, et al., 2006] depending on the type of preferences; define all sections with strictly monotonic preferences between the extremum platforms;
   (b) in the case of monotonic preferences, the entire one-dimensional set of prizes is one section;
   (c) elicit z number of knots of the local utility function in each section using PE, CE, LE, UE, etc.; arctg-approximate the utility function following the procedures from [Tenekedjiev, et al., 2007], and if it is of poor quality - replace it by linear interpolation;
   (d) in the case of non-monotonic preferences: construct the global utility function by using the algorithms from [Nikolova, et al., 2006] several times; in the case of monotonic preferences the local utility function coincides with the global one.
5. Elicit the uncertainty intervals of the scaling constants.
6. Use the numerical [Tenekedjiev, 2008] or the simulation [Nikolova, 2007a] realization of the uniform method to analyze the sum of the scaling constants and to find their point estimates.
7. Construct the multi-dimensional utility function in a form defined by (5).

5 Summary and Conclusions

The construction of a multi-dimensional utility function was described as a problem that depends on the complexity of the situation, and the type of preferences of the FRDM. The theoretical approaches that apply to each stage were summarized in the paper. On that basis, the last section presented a detailed algorithm that can be applied to construct the utility function of a FRDM over multi-dimensional prizes. The algorithm adopted the suggestion that from a practical point of view, one should try to establish mutual utility independence between (groups of) attributes, elicit scaling constants, and depending on their sum define the final additive or multiplicative form of the multi-dimensional utility function.

Future research in that respect would focus on verification of the elaborated algorithm, i.e. real-life decision problems should be identified where prizes are described by multi-dimensional vectors, and where different types of preferences (monotonic, non-monotonic) exist over the attributes, so that all stages of the algorithm could be applied.

References

[1] R. Clemen, *Making Hard Decisions: an Introduction to Decision Analysis*, Second Edition. Duxbury Press, Wadsworth Publishing Company, 1996.

[2] Y. Engel, M. Wellman, CUI Networks: A Graphical Representation for Conditional Utility Independence, *Proc. 21 National Conference on Artificial Intelligence*, pp. 1137-1142, 2006.

[3] P.H. Farquhar, Utility Assessment Methods, *Management Science*, Vol. 30, No. 11, pp. 1283-1300, 1984.

[4] G. E. Forsythe, A. Malcolm, C. B. Moler, *Computer Methods for Mathematical Computations*, Prentice Hall, 1977.

[5] S. French, *Decision Theory: an Introduction to the Mathematics of Rationality*, Ellis Horwood, 1993.

[6] S. French, D.R. Insua, *Statistical Decision Theory*, Arnold, London, 2000.

[7] R. L. Keeney, H. Raiffa, *Decisions with Multiple Objectives: Preference and Value Tradeoffs*, Cambridge University Press, 1993.

[8] J. Kiefer, Sequential Minimax Search for a Maximum, *Proc. American Mathematical Society*, Vol. 4, pp. 502-506, 1953.

[9] M. McCord, R. De Neufville, *Lottery Equivalents': Reduction of the Certainty Effect Problem in Utility Assessment*, Management Science, Vol. 32, pp. 56-60, 1986.

[10] N. D. Nikolova, A. Shulus, D. Toneva, K. Tenekedjiev, Fuzzy Rationality in Quantitative Decision Analysis, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 9, No. 1, pp. 65-69, 2005.

[11] N. D. Nikolova, K. Hirota, C. Kobashikawa, K. Tenekedjiev, Elicitation of Non-Monotonic Preferences of a Fuzzy Rational Decision Maker, *Information Technologies and Control*, Year IV, Vol. 1, pp. 36-50, 2006.

[12] N. D. Nikolova, Uniform Method for Estimation of Interval Scaling Constants, *Engineering and Automation Problems*, Vol. 1, pp. 79-90, 2007a.

[13] N. D. Nikolova, Empirical Connection between the Number of Elicited Knots and the Quality of Analytical Approximation of a One-Dimensional Utility Function, *Machine Building and Machine Learning, Economics and Management Series*, Year II, Book 1, pp. 99-111, 2007b (in Bulgarian)

[14] J. W. Pratt, Risk Aversion in the Small and in the Large, *Econometrica*, Vol. 32, pp. 122-136, 1964.

[15] W. H. Press, S. A. Teukolski, W.T. Vetterling, B.P. Flannery, *Numerical Recipes - the Art of Scientific Computing*, Cambridge University Press, 1992.

[16] S. Stoianov, *Optimization of Technological Processes*, Tehnika, 1993 (in Bulgarian).

[17] K. Tenekedjiev, N.D. Nikolova, D. Dimitrakiev, Application of the Triple Bisection Method for Extraction of Subjective Utility Information, *Proc. Second International Conference "Management and Engineering'2004"*, Vol. 2, No. 70, pp. 115-117, 2004.

[18] K. Tenekedjiev, N.D. Nikolova, R. Pfliegl, Utility Elicitation with the Uncertain Equivalence Method, *Comptes Rendus De L'Academie Bulgare des Sciences*, Vol. 59, Book 3, pp. 283-288, 2006.

[19] K. Tenekedjiev, N. D. Nikolova, D. Dimitrakiev, Analytical One-Dimensional Utility - Comparison of Power and Arctg-Approximation, *Engineering Science*, Vol. 4, pp. 19-32, 2007 (in print)

[20] K. Tenekedjiev, Justification and Numerical Realization of the Uniform Method for Finding Point Estimates of Interval Elicited Scaling Constants, *Fuzzy Optimization and Decision Making*, 2008 (in print)

[21] J. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Second Edition, Princeton University Press, 1947.

[22] P. Wakker, D. Deneffe, Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown, *Management Science*, Vol. 42, pp. 1131-1150, 1996.

Natalia Nikolova, Sevda Ahmed, Kiril Tenekedjiev
Technical University - Varna
Dept. Economics and Management
1 Studentska Str., 9010 Varna, Bulgaria
E-mail: natalia@dilogos.com, sevdadan@hotmail.com, kiril@dilogos.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 427-432

Spi Calculus Analysis of Otway-Rees Protocol

Horea Oros, Florian Boian

Abstract: In this paper we present the Otway-Rees authenticated key transport protocol. We present informally the Spi calculus, an extension of π-calculus with cryptographic primitives, and use it to express and analyze the Otway-Rees protocol.
Keywords: key transport protocol, spi-calculus, formal analysis

1 Introduction

There are many cryptographic protocols that were used for a long time when attacks against them were discovered [6],[11]. The existence of these attacks, that show up long after the protocols are created, demonstrates that it is quite difficult to devise protocols that posses the needed security properties. What is needed is a theory that enables us to describe and reason about cryptographic protocols, that could formally prove that a protocol indeed has the stated security properties or, better yet, that could find security flaws in a protocol.

The theories for formal reasoning about cryptographic protocols, can be divided in two major approaches: a computational approach, in proving the security of cryptographic protocols, initiated by Bellare and Rogaway [2] and a symbolic approach that relies on simple but effective formal language approach and includes many techniques (theorem proving, model checking, BAN logic [4] and extensions of it, the NRL protocol analyzer [7], the CSP approach [5]).

So, there are many notations for describing security protocols some of which have been long used in security literature. The main shortcoming of these notations is that they do not provide a precise and solid basis for reasoning about protocols. Some notations have a fairly clear connection to the intended implementation whereas some are more formal the relationship with the implementation being more subtle. One notations that we can use for describing security protocols is the Spi Calculus. The Spi calculus appears to be a theory that yields more convincing proofs of correctness for security protocols, such as those for authentication and for electronic commerce. The Spi calculus belongs to the symbolic approach.

Spi calculus is an extension of Milner's π calculus, designed for the description and analysis of cryptographic protocols. It is particularly well suited for studying authentication protocols, protocols designed for electronic commerce etc. While the π calculus suffices for some abstract protocols, the spi calculus enables us to treat the issues involved in cryptographic protocols in more detail.

In this paper we use Spi calculus to express the Otway-Reese protocol providing a foundation for further studying of the subject.

The remainder of this paper is organized as follows: section 2 presents the Otway-Rees protocol that provides authenticated key transport without providing entity authentication or key confirmation. The notation and presentation of the protocol is an informal one - we present the system setup, protocol messages and protocol actions. Section 3 presents the syntax and semantics of one version of the Spi calculus (there are many others). In section 4 we express the Otway-Reese in Spi calculus. The concluding section will close the presentation.

2 Otway-Rees Protocol

The Otway-Rees protocol [10] is a server-based protocol providing authenticated key transport (with key authentication and key freshness assurance) in only 4 messages without requiring timestamps as it is in Kerberos. It does not, however provide entity authentication or key confirmation.

In this paper A and B are two general users, T is the server or trusted third party and I is an attacker or intruder. K_{AB} is the shared between A and B.

We present in informal notation the Otway-Rees protocol.

Otway-Rees Protocol:
SUMMARY: B interacts with trusted server T and party A.
RESULT: establishment of fresh shared secret key K between A and B.

1. *Notation.* $\{M\}_K$ encryption of message M with key K using a symmetric encryption algorithm. For security reasons the encryption function has a built-in data integrity mechanism to detect message modification. K_{AB}

is a session key generated by T for A and B to share. N_A and N_B are nonces chosen by A and B, respectively, to allow verification of key freshness (thereby detecting replay). M is a second nonce chosen by A which serves as a transaction identifier.

2. *System setup*. T shares symmetric keys K_{AT} and K_{BT} with A, B, repectively.

3. *Protocol messages.*

$$\begin{array}{lll} \text{Message 1}: & A \to B: & M, A, B, \{N_A, M, A, B\}_{K_{AT}} \\ \text{Message 2}: & B \to T: & M, A, B, \{N_A, M, A, B\}_{K_{AT}}, \{N_B, M, A, B\}_{K_{BT}} \\ \text{Message 3}: & T \to B: & M, \{N_A, K_{AB}\}_{K_{AT}}, \{N_B, K_{AB}\}_{K_{BT}} \\ \text{Message 4}: & B \to A: & M, \{N_A, K_{AB}\}_{K_{AT}} \end{array}$$

4. *Protocol actions*. Perform the following steps each time a shared key is required.

   (a) A encrypts data for the server containing two nonces, N_A and M, and the identities of itself and the party B to whom it wishes the server to distribute a key. A sends this and some plaintext to B in message 1.

   (b) B creates its own nonce N_B and an analogous encrypted message (with the same M), and sends this along with A's message to T in message 2.

   (c) T uses the cleartext identifiers in message 2 to retrieve K_{AT} and K_{BT}, then verifies the cleartext (M, A, B) matches that recovered upon decrypting both parts of message 2. (Verifying M confirms that the encrypted parts are linked.) If so, T inserts a new key K_{AB} and the respective nonces into distinct messages encrypted for A and B, and sends both to B in message 3.

   (d) B decrypts the third part of message 3, checks N_B matches that sent in message 2, and if so passes the first part on to A in message 4.

   (e) A decrypts message 4 and checks N_A matches that sent in message 1.

   If all checks pass, each of A and B are assured that K_{AB} is fresh (due to their respective nonces). A knows that B is active as verification of message 4 implies B sent message 2 recently; B however has no assurance that A is active until subsequently use of K_{AB} by A, since B cannot determine if message 1 is fresh.

3 The Spi Calculus

Abadi and Gordon have introduced the Spi calculus in [1] as an extension of Milner's π-calculus [8] with cryptographic primitives. The Spi calculus is designed for the description and analysis of cryptographic protocols. These protocols rely on cryptography and on communication channels with properties like authenticity and privacy. Accordingly, cryptographic operations and communication through channels are the main ingredients of the spi calculus.

There are many versions of π-calculus and Spi calculus. We review the syntax and semantics of a particular version of the Spi calculus with primitives for shared key cryptography, hashing, public-key encryption and digital signatures. The differences with other versions is not our concern here.

We assume an infinite set of *names*, to be used for communication channels and an infinite set of *variables*. We let *m, n, p, q* and *r* range over names, and let *x, y,* and *z* range over variables.

The set of *terms* is defined by the grammar:

L, M, N::=	terms
n	name
(M, N)	pair
0	zero
$suc(M)$	successor
x	variable
$\{M\}_N$	shared-key encryption (N is the key)
$H(M)$	hashing
M^+	public part, (public half of M)
M^-	private part, (private half of M)
$\{[M]\}_N$	public-key encryption
$[\{M\}]_N$	private-key signature

The set of *processes* is defined by the grammar:

$P, Q, R ::=$	processes
$\overline{M}\langle N\rangle.P$	output
$M(x).P$	input
$P\|Q$	composition
$(\nu n)P$	restriction
$!P$	replication
$[M \textit{ is } N]P$	match
0	nil
$let(x,y) = M \textit{ in } P$	pair splitting
$case\ M\ of\ 0 :\ P\ suc(x) :\ Q$	integer case
$case\ L\ of\ \{x\}_N \textit{ in } P$	shared key decryption
$case\ L\ of\ \{[x]\}_N \textit{ in } P$	decryption with private key
$case\ N\ of\ [\{x\}]_M \textit{ in } P$	signature check with public key

Intuitively, the constructs of the Spi calculus have the following meanings:

- The basic computational step and synchronization mechanism in the π-calculus is *interaction*, in which a term N is communicated from an output process to an input process via a named channel, m.
  - An *output process* $\overline{m}\langle N\rangle.P$ is ready to output on channel m. If an interaction occurs, term N is communicated on m and then process P runs.
  - An *input process* $m(x).P$ is ready to input from channel m, then process $P[N/x]$ runs ($P[N/x]$ means term N replaces each free occurrence of variable x in process P). The variable x is bound in process P. (The general forms $\overline{M}\langle N\rangle.P$ and $M(x).P$ of output and input allow for the channel to be an arbitrary term M. The only useful cases are for M to be a name, or a variable that gets instantiated to a name.)
- a *composition* $P|Q$ behaves as processes P and Q running in parallel. Each may interact with the other on channels known to both, or with the outside world, independently of the other.
- A *restriction* $(\nu n)P$ is a process that makes a new, private name n, which may occur in P, and then behaves as P. The name n is bound in P.
- A *replication* $!P$ behaves as an infinite number of copies of P running in parallel.
- A *match* $[M \textit{ is } N]P$ behaves as P provided that terms M and N are the same; otherwise it is stuck, that is, it does nothing.
- The *nil* process **0** does nothing.
- A *pair splitting* $let(x,y) = M \textit{ in } P$ behaves as $P[N/x][L/y]$ if term M is the pair (N,L), and otherwise it is stuck. The variables x and y are bound in P.
- An *integer case* process $case\ M\ of\ 0 :\ P\ suc(x) :\ Q$ behaves as P if term M is 0, as $Q[N/x]$ if M is $suc(N)$, and otherwise is stuck. The variable x is bound in the second branch Q.
- The process $case\ L\ of\ \{x\}_N \textit{ in } P$ attempts to decrypt the term L with key N. If L is a ciphertext of the form $\{M\}_N$, then the process behaves as $P[M/x]$. Otherwise the process is stuck. The variable x is bound in P.
- $H(M)$ represents the hash of message M. The absence of a construct for recovering M from $H(M)$ corresponds to the assumption that H cannot be inverted. The lack of any equations $H(M) = H(M')$ correspond to the assumption that H is collision free.
- The process $case\ L\ of\ \{[x]\}_N \textit{ in } P$ decrypts with the private part of N message L where L is the encryption of x with the public part of N; $L = \{[M]\}^+_N$. The variable x is bound in process P.
- The term $[\{M\}]_N$ represents the result of the signature of M with N. The variable x is bound in P in the construct $case\ N\ of\ [\{x\}]_M \textit{ in } P$. This construct is dual to $case\ L\ of\ \{[x]\}_N \textit{ in } P$. The new construct is useful when N is a public key K^-; then it binds x to the M such that $[\{M\}]^-_K$ is L, if such an M exists. We are assuming that M can be recovered from the result of signing it.

Some standard and significant assumptions are implicit above:

- The only way to decrypt an encrypted packet is to know the corresponding key.
- An encrypted packet does not reveal the key that was used to encrypt it.
- There is sufficient redundancy in messages so that the decryption algorithm can detect weather a cipher text was encrypted with the expected key.

4 Otway-Rees protocol in Spi

Now, after we have presented the Otway-Rees protocol and the Spi calculus we will express the protocol in Spi. We will allow for one principal to be involved in more than one instance of the protocol even simultaneously.

We consider a system with a server T and n other principals. Each of these principals are allowed to initiate instances of the protocol and are willing to participate in any instance of the protocol initiated by other entities. For the names of the principals we use the terms $suc(0)$, $suc(suc(0))$, ..., which we abbreviate to $\underline{1}$, $\underline{2}$, ... We assume that each principal has an input channel; these input channels are public and have the names c_1, c_2, ..., c_n and c_T for the server.

In our Spi calculus representation, we use several convenient abbreviations. We rely on pair splitting on input and on decryption:

$$\begin{array}{rcl} c(x_1,x_2).P & \triangleq & c(y).let(x_1,x_2) = y \text{ in } P \\ case\ L\ of\ \{x_1,x_2\}_N \text{ in } P & \triangleq & case\ L\ of\ \{y\}_N \text{ in } let(x_1,x_2) = y \text{ in } P \end{array}$$

This syntax for pairs and pair splitting can be generalized to allow arbitrary tuples. This is useful because the protocol messages consists of many components. We use the following standard abbreviations, given inductively for any $k \geq 2$:

$$\begin{array}{rcl} (N_1,\ldots,N_k,N_{k+1}) & \triangleq & ((N_1,\ldots,N_k),N_{k+1}) \\ let(x_1,\ldots,x_k,x_{k+1}) = N \text{ in } P & \triangleq & let(y,x_{k+1}) = N \text{ in } let(x_1,\ldots,x_k) = y \text{ in } P \end{array}$$

In Spi calculus we represent the nonces of the protocol as newly created names.

For the composition of a finite set of processes we use the standard notation:

$$\prod_{i\in 1..k} P_i \triangleq P_1 | \ldots | P_k$$

In an encrypted pair of the form $\{(N,N')\}_{N''}$, we omit the inner brackets, as in informal notation.

Informally, an instance of the protocol is determined by a choice of parties (who is A and who is B) and by the message that is sent after key establishment. An instance I is a triple (i,j,m) such that i is the initiator of an instance of protocol and j is the second principal involved in the protocol. We assume that there is an abstraction F representing the behaviour of any principal after receiving the last message of the protocol.

Given an instance (i,j,m), the following process corresponds to the role of A:

$$\begin{array}{rcl} A(i,j,m) & \triangleq & (\nu N_A)\overline{c_j}\langle(M,\underline{i},\underline{j},\{N_A,M,\underline{i},\underline{j}\}_{K_{AT}})\rangle \,|\, c_i(x_M,x_{cipher}). \\ & & .[x_M \text{ is } M] case\ x_{cipher}\ of\ \{x_{nonce},x_{key}\}_{K_{iT}} \text{ in} \\ & & [x_{nonce} \text{ is } N_A]\overline{c_j}\langle\{m\}_{x_{key}}\rangle \end{array}$$

A creates fresh nonce that is restricted to the current process N_A, sends on the input channel of $\underline{j}$ the message consisting of an instance identifier M, its identity $\underline{i}$, the identity of the recipient $\underline{j}$ and a message encrypted with the key that she shares with the server. The encrypted message consists of the nonce N_A, the protocol instance identifier M and the identity of the initiator of the protocol and the identity of the intended recipient.

The following process corresponds to the role of B for principal j:

$$\begin{aligned} B(j) \triangleq\ & c_j(y_M, y_A, y_B, y_{cipher}).[y_B \text{ is } j](\nu N_B) \\ & \overline{c_T}\langle (y_M, y_A, \underline{j}, y_{cipher}, \{N_B, y_M, y_A, \underline{j}\}_{K_{jT}})\rangle \mid \\ & \mid c_j(y'_M, y'_{cipher}, y''_{cipher}).[y_M \text{ is } y'_M] case\ y''_{cipher}\ of \\ & \{y_{nonce}, y_{key}\}_{K_{jT}}\ in\ [y_{nonce} \text{ is } N_B]\overline{c_i}\langle y'_{cipher}\rangle. \\ & .c_j(m_{cipher}).case\ m_{cipher}\ of\ \{y_m\}_{y_{key}}\ inF(y_m) \end{aligned}$$

Process B waits for a message on its input channel that consists of four parts: y_M a protocol identifier, y_A the identity of the principal who pretends that created the message and a cipher text. Next, process B verifies that the third part of the message he received corresponds to his identity. After that he generates a fresh nonce N_B that is restricted to this process and sends to server T along his input channel c_T a message consisting of five parts: the protocol instance identifier y_M, the identity of A y_A, the cipher text that he received from A y_{cipher} (these three parts were received in the previous step), his identity $\underline{j}$ and a cipher text encrypted with the key that B shares with the server. The encrypted part consists of jour parts: a nonce N_B generated by B, the protocol instance identifier y_M, the identity of A y_A and his identity $\underline{j}$. In parallel B waits for a message from the server that consists of three parts: the protocol instance identifier y'_M and two cipher texts y'_{cipher} and y''_{cipher}. The protocol instance identifier is compared to the one received previously and if they match the process continues with the decryption of the second cipher text, otherwise the process is stuck. The second cipher text is decrypted with the key that B shares with the server and obtains the nonce $y_n once$ and the key y_{key} generated by the server for A and B to share. The nonce is compared to the one generated by B N_B and if they match the first cipher text y'_{cipher} is sent to A on its input channel c_i. Otherwise, the process is stuck. After that process B may communicate securely on public channels c_i and c_j with process A using the shared key y_{key}.

The behaviour of the server T is expressed with the following Spi calculus process:

$$\begin{aligned} T \triangleq\ & c_S(z_M, z_A, z_B, z'_{cipher}, z''_{cipher}).\prod_{i\in 1..n}[z_A \text{ is } \underline{i}]\prod_{j\in 1..n}[z_B \text{ is } \underline{j}] \\ & case\ z'_{cipher}\ of\ \{z^A_{nonce}, z'_M, z'_A, z'_B\}_{K_{iT}}\ in\ [z_M \text{ is } z'_M][z_A \text{ is } z'_A][z_B \text{ is } z'_B] \\ & case\ z''_{cipher}\ of\ \{z^B_{nonce}, z''_M, z''_A, z''_B, \}_{K_{jT}}\ in\ [z_M \text{ is } z''_M][z_A \text{ is } z''_A][z_B \text{ is } z''_B] \\ & (\nu k)\overline{c_j}\langle (z_M, \{z^A_{nonce}, k\}_{K_{iT}}, \{z^B_{nonce}, k\}_{K_{jT}})\rangle.0 \end{aligned}$$

The server waits for a message on its input channel c_S, message that consists of five parts: the first part is the protocol instance identifier z_M, the second and the third are the identities of the two principals that wish to communicate securely z_A, z_B and the last two parts are two cipher texts encrypted with the keys that the server shares with A and B, respectively. The server decrypts the two parts and verifies if the protocol instance identifier that is encrypted in both messages matches z_M. Also, the server verifies if the identities of the two entities that are encrypted matches the identities that are sent in plain text (z_A and z_B). If all these verification succeed the server generates a new key k that is restricted to this process and sends to B on its input channel c_j a message consisting of three parts: the protocol instance identifier and two encrypted messages. The two encrypted messages consists of the key k and the nonce generated by A and B respectively. The keys used for encryption are the ones that the server shares with A and B, respectively. Thus, the server completes his role in the protocol instance identified by z_M.

Finally we define a whole system, parameterized on a list of instances of the protocol:

$$\begin{aligned} OR(I_1, \ldots, I_m) \triangleq\ & (\nu K_{1S}), \ldots, (\nu K_{nS}) \\ & (A(I_1)) \mid, \ldots, (A(I_m)) \mid !S \mid !B(1) \mid, \ldots, \mid !B(n)) \end{aligned}$$

The expression $OR(I_1, \ldots, I_m)$ represents a system with m instances of the protocol. The server is replicated. The replication of the B processes means that each principal is willing to play the role of receiver in any number of runs of the protocol in parallel. Thus, any two runs of the protocol can be simultaneous, even if they involve the same principals.

5 Conclusion

We have presented the Otway-Rees protocol that is a server-based protocol providing authenticated key transport (with key authentication and key freshness assurances) in only 4 messages, but does not provide entity authentication or key confirmation.

We presented the Spi calculus, an extension of Milner's π-calculus with cryptographic primitives. The Spi calculus can be used in the description and formal analysis of cryptographic protocols.

In section 4 we expressed the Otway-Rees protocol in Spi calculus.

References

[1] Martin Abadi, D. Gordon, "A Calculus for Cryptographic Protocols. The Spi Calculus", Technical Report No. UCAM-CL-TR-414, University of Cambridge, Computer Laboratory, 1996.

[2] Mihir Bellare, Phillip Rogaway, "Entity authentication and key distribution", *Proc. CRYPTO 93*, ed. Douglas R. Stinson, *Lecture Notes in Computer Science* No. 773, Springer, pp. 232–249, 1994.

[3] Colin Boyd, Wenbo Mao, "On a Limitation of BAN Logic", *Advances in Cryptology-EUROCRYPT'93. Lecture Notes in Computer Science*, No 765, Ed. Tor Helleseth, pp. 240–247, 1993.

[4] Michael Burrows, Martin Abadi, Roger Needham, "A logic of authentication", *Proceedings of the Royal Society of London*, Volume A426, pp. 233–271, 1989.

[5] C.A.R. Hoare, "Communicating Sequential Processes", *Communications of the ACM*, Volume 21, No. 8, Series in Computer Science, 1978.

[6] G. Lowe, "An Attack on the Needham-Schroeder Public-Key Authentication Procotol", Information Processing Letters, Volume 56, No. 3, pp. 131–133, 1995.

[7] Cathrin Meadows, 'The NRL Protocol Analyzer: An Overview", *Journal of Logic Programming*, Volume 26, No. 2, pp. 113–131, 1996.

[8] Robin Milner, J. Parrow, D. Walker, "A calculus of mobile processes, parts I and II", *Information and Computation*, pp. 1–40, 41–77, 1992.

[9] Robin Milner, "Communicating and mobile systems: the π-calculus", *Cambridge Universtiy Press*, ISBN 0 521 64320 1, 1999.

[10] Dave Otway, Owen Rees, "Efficient and Timely Mutual Authentication", *ACM Operating System Review*, 21(1) January, pp. 8–10, 1987.

[11] Guilin Wang, Sihan Qing, "Two New Attacks Against Otway-Rees Protocol", *IFIP/SEC2000, Information Security*, Vol. 16th World Computer Congress 2000, International Academic Publishers, pp. 137–139, 2000.

[12] "Ding Yi-Qiang", "The Formal Analysis of Cryptographic Protocols", PhD Theses, Institute of Software, The Chinese Academy of Science, June 1999.

Horea Oros
University of Oradea
Faculty of Sciences
Department of Mathematics and Computer Science
410087, Universitatii St., 1, Oradea, Romania
E-mail: horos@uoradea.ro

Florian Boian
Babes-Bolyai University of Cluj-Napoca
Faculty of Mathematics and Computer Science
Department of Computer Science
400084, M. Kogalniceanu St., 1, Cluj-Napoca, Romania
E-mail: florin@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 433-436

Technology to Support Education Software Solutions for Quality Assurance in e-learning

Iulian Pah, Constantin Oprean, Ioana Moisil, Claudiu Kifor

Abstract: Quality assurance in e-learning is of growing interest because of today proliferation of e-learning products, most addressed to continuing education, distance learning, but also for students in schools and universities. In our paper we are presenting some results of a couple of research projects that had as main objective to achieve a better quality management of all university's aspects, including e-learning, through an advanced web-based, multi-agent, knowledge management system. The proposed solutions have advanced functions for extracting the quality indicators from the university data base, online analysis of indicators' values and for recommending the suitable measures in order to adjust inappropriate values of individual indicators. The first project is *eUNIV* - an e-business solution of knowledge management for the academic environment. The second one is *e-EdU-Quality* a project that developed a quality system starting from the experience obtained in the *eUNIV* project, and based on the university information system, the pilot study being implemented on the university intranet. Both *eUniv* and *e-EdUQuality* projects have been developed by the Lucian Blaga University of Sibiu. The third project is a conceptual model for e-learning systems - the DANTE project - Socio-Cultural Models implemented through Multi-Agent Architecture for eLearning - based on a global model for the virtual education environment, student centred, that facilitates the learning through collaboration as a form of social interaction.
Keywords: quality of education, higher education, quality assurance, e-learning, levels of e-learning, quality systems, knowledge management, multi-agent system.

1 Introduction

Quality assurance in e-learning is of growing interest. This is due mainly because of today proliferation of e-learning products, most addressed to continuing education, distance learning, and also to students in schools and universities. Another reason of the raising interest in quality assurance (QA) is that e-learning is the main tool of internationalization of education. By quality assurance we understand the means through which an institution ensures and confirms that there are suitable conditions for students to achieve the standards set by it or by another awarding body. At the European level [2], we can underline different stages for quality assurance and quality management in education expressed by the following declarations and conventions: Sorbonne declaration (1998) referring to the European Space for Higher Education; Bologna declaration (1999) is a pledge by 29 countries to reform the structures of their higher education systems in a convergent way until 2010 by promoting European cooperation for quality assurance; Lisbon convention (2000), pinpoints the idea of a competitive economic society based on knowledge; Salamanca convention (2001); Berlin declaration (2003), where is established that " the main responsibility for quality assurance in higher education belongs to each institution"; Bergen convention (2005), a set of recommendations for guides, standards and procedures, national and international framework for qualifications, minimum number of ECTS, life long learning.

At the global level also there is an interest for academic quality assurance. For example, UNESCO and OECD, are involved to elaborate and adoption some "orientations" regarding the quality of services offered by transnational institutions. The problem is to impose the quality not only in the national institutions but also in the transnational one. The interests for quality and e-learning technologies can be observed also in the national initiatives like e-Fit Austria [3, 4]. Some of the main objectives of this program includes: easy access to innovative services, high quality content for education, science, training and culture all this using the information technology to create a better and more efficient service for the educational system.

The relations between the European framework and national frameworks are based on transparency, visibility and comparability, therefore any higher education institute is responsible to develop a culture for quality which means politics, techniques and practices consequently applied and documented to obtain those results/performances that are in concordance with the proposed objectives. At European level and also at international level there are organizations concerned with the aspects of quality assurance in higher education: European Association for Quality Assurance in Higher Education (ENQA), the Nordic Quality Assurance Network in Higher Education (NOQA), the Network Agencies for Higher Education Quality Assurance from Central and Oriental Europe, the

Network D-A-CH - created to prepare reciprocal recognition of credit decision taken by Austrian, German and Swiss Councils of Accreditation. The Romanian Ministry of Education, Research and Youth has issued a regulation that says that: "each higher education institute from Romania (...) is the main responsible for the quality of educational services and also for quality assurance ..., having the obligation that starting with the 2005/2006 academic year to implement and use its own assurance quality system" [2]. Today, though many universities declare that they have a quality system, the most of them have this system only on paper and despite the fact that they have a quality manual and procedures, the functionality of these systems is very poor.

In order to achieve the goals set for 2010 by the Bologna Process national education bodies called upon universities to establish performance indicators to measure progress towards these goals.

At the level of a university defining a quality assurance system starts by determined an agreed set of criteria, standards and models for quality assurance and in identifying the main critical success factors for self assessment. The stakeholders in the quality process have to be identified, as their values will determine how quality itself is defined and measured. In general there are four stakeholder groups: policy-makers and their administration, educational entities providing programs or courses, teachers or trainers and students or trainees. Each of these groups has different interests leading to different perspectives on quality. For example, effective teaching as judged by a student may not be considered cost efficient in economic terms. The main facets of quality: *quality control measures*, implemented at the level of the university, standards and procedures for *quality assurance and total quality management* - TQM (customer focus; continuous improvement; quality assurance of internal processes; process orientation; and prevention instead of inspection) have to be considered. The objectives of quality assurance are articulated in order to clarify the purpose of the pursuit and the indicators of quality defined. Most of the TQM indicators are relayed to the outcomes of education. The other indicators relate to control of resources, and of the educational content.

The purpose of indicators is twofold: they provide information to policy-makers to assist in policy formulation, and they demonstrate accountability. Specialists in quality assessment consider that indicators ought to be developed so that they are: policy-relevant, user-friendly (timely, comprehensible and few in number), derived from context, valid and reliable, and last but not least measurable at a reasonable cost. The European Commission Working Committee on Quality Indicators identified four main groups of quality indicators for the quality of education: attainment indicators, success and transition indicators, monitoring of school education indicators and resources and structures indicators (EC 2000).

Having specified and approved the group of indicators at the level of the university, we started several research projects in order to develop intelligent software tools to assist education quality audits and to make operational the ongoing process of university performance evaluation according to agreed measures. In the followings we will present the main components of the system as a result of three research projects. The general architecture of the system is presented in figure 2.

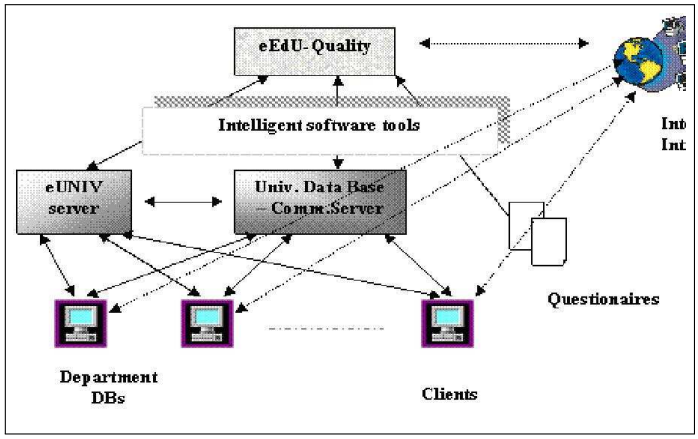


Figure 1: eEdu-Quality architecture

2 The eUNIV and e-EdU-Quality Projects

The *eUNIV* project was aimed to re-engineering an educational organization, based on a business solution - eyeKNOW - developed for knowledge management in an enterprise by our industrial partner, *Wittmann and Partner Computer Systems*. The functionality of eyeKNOW has been extended to fit the requirements of an academic environment. The result was eUNIV, a client-server, Lotus Notes Domino application. The pilot site was a department of the Faculty of Engineering. The first evaluation results have shown that the solution is a feasible one. That means that strategies applied to optimize the overall activities in a commercial organization can be successfully applied, after customization, to an educational environment. eUNIV is centred on the concept of project. A project is defined as a set of activities and tasks oriented toward a goal. Resources are allocated to each project. Some projects can share the same resources. The eUNIV system enables not only a better management of all kinds of documents and projects, but it is an environment that allows educational staff to adapt to a new style of work: to share resources, projects, to cooperate without frontiers, in an organized and structured way. The pitfalls of the system are those of all attempts to standardization.

Each project has a coordinator, a team and resources. For each project there is an agenda, visible by all academic and non-academic staff, a mail-box, a special mail-box for students and a chat space, where colleagues can start less formal discussions.

The access rights are allocated by the system administrator. There are four categories of access rights: public - anyone can access the information from the web or from a workstation, priority 0 (administrator and head of department), priority 1 gives access to the projects coordinators (access granted to all system information, but some are read-only), priority 2 gives access only to the project information and the ones that are public, of course and priority 3, only to administrative and public information. After the first evaluation it is expected that we are going to make the access more flexible, but for the moment we are keeping it like this.

The *e-EdU-Quality* system is enriching the eUNIV application with functionalities that allow to extract the quality indicators from the university data base, online analysis of indicators' values and to recommend the suitable measures in order to correct, if this is the case, inappropriate values of individual indicators. Several intelligent software tools that assist the process of quality assurance and management have been developed and are in the process of implementation. The main tools are: students' performance indicators extractor, electronic voting for the selection of grant's proposals; quality evaluation questionnaires manager (questionnaires generator, distributor and analyzer); resources management indicators. Most of the developed tools are based on multi-agent technology. The e-EdU-Quality multi-agent system infrastructure uses different kinds of agents: *interface agents* that interact with users, receive user input, and display results, *task agents*, that help users perform tasks, formulate problem-solving plans and carry out these plans by coordinating and exchanging information with other software agents, *information agents*, providing intelligent access to heterogeneous collections of information sources, and also *middle agents*, acting as intermediary between agents that request services and agents that provide services.

3 The DANTE Project

DANTE *Socio-Cultural Models implemented through multi-agent architecture for e-learning* has as main objective the development of a global model for the virtual education system, student centred, that facilitates the learning through collaboration as a form of social interaction. In our vision, the global model requires its own universe in which the human agents interact with software agents. In the virtual worlds of software agents, things must be similar with what is happening in a real world, and this is visible if we look at the metaphor "computing as interaction" or at the "emergent synthesis" design methodology.

The global model is considered the core of an e-learning system. From a pedagogical point of view, DANTE is combining the hierarchical way of learning with the collaborative one. The proposed *e-Learning* system has a general architecture with three levels: user, intermediary, supplier educational space, on each level heterogeneous families of human and software agents are interacting. The main human actors are: the *student*, the *teacher* and the *tutor*. In the virtual learning environment we have the corresponding agents.The human actors are interacting with the e-learning system via several agentified environments. The **teacher** (human agent) is assisted by two types of software agents: *personal assistant* (classic interface agent) and **didactic** *assistant*. The **SOCIAL** *agentified environment* has social agents and a database with *group models* (profiles of social behaviour). The *agentified DIDACTIC environment* assists the cognitive activities of the student and/or of the teachers. The student (human agent) evolves in an agentified environment with three types of agents. He/she has a personal assistant (software interface agent) who monitors all the student's actions and communicates (interacts) with all the other agents,

with the agentified environments of other students and with the teacher's agentified environment. The student has at his/her disposal two more agents: the **TUTOR** and the *mediating agent*.The **TUTOR** assistant evaluates the educational objectives of the student and recommends her/him some kind of activities. The decisions are based on the knowledge of the students' cognitive profile (which takes into account the social component). The **TUTOR** agent interacts with the personal assistant of the student, with the mediating agent and with the social agentified environment. As the system is conceived, the accent is put on collaboration activities between students, which consist in knowledge exchange, realization of common projects, tasks' negotiation, sharing resources, common effort for the understanding of a subject, problem-solving in-group.

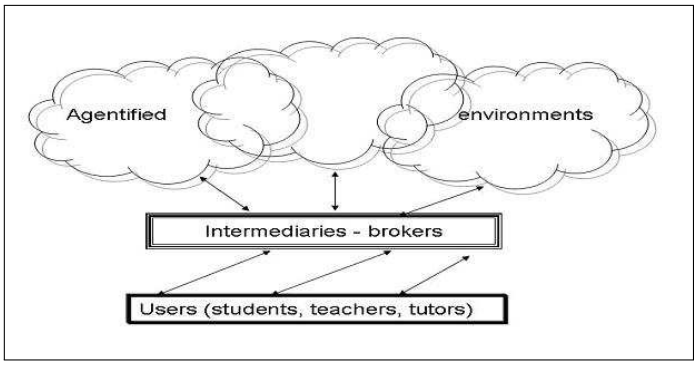


Figure 2: Three layered architecture of the DANTE system

4 Conclusions

In our paper we have presented an approach to quality assurance in higher education based on a set of software tools designed to assist the evaluation process and to ensure that interdependent processes are properly coordinated in order to comply with the university's strategy and to achieve the predetermined goals. The system is under development, a pilot being tested at the faculty of engineering (quality of e-learning master degree programs) and at the level of the department of research (e-voting for the selection of grant proposals). The first results are promising, but further tests must be carried on. Quality assurance is a continuing process and must be the university number one priority. To quote François Tavenas "The only standard that should guide universities and all their players is that of quality judged with reference to best practice in international university circles".

References

[1] P. Dawson, and G. Palmer, (1995). *Quality management*. Australia: Longman, 1995.

[2] Nicolae Dragulanescu, Motivatii si obstacole ale asigurarii calitatii în învatamântul superior, Fundatia Româna pentru Promovarea Calitatii, Bucuresti, 2005.

[3] Austrian Federal Ministry of Education, Science and Culture, Austrian Education News, 43/2005.

[4] www.eFit.at

[5] D. Garvin, (1988). *Managing quality*. New York: Macmillan, 1988.

[6] A. Hope, "Quality Assurance", *The Development of Virtual Education: A Global Perspective* (ed. Glen Farrell), Commonwealth of Learning, Vancouver, Canada, pp 125-140, 2001.

Iulian Pah
"Babes-Bolyai" University, Cluj-Napoca, Romania

Constantin Oprean, Ioana Moisil, Claudiu Kifor
"Lucian Blaga" University of Sibiu, Romania
4, Emil Cioran Str., Sibiu 550025, Romania

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 437-441

Multi-Level Database Mining Using AFOPT Data Structure and Adaptive Support Constrains

Mirela Pater, Daniela E. Popescu

Abstract: Finding frequent itemsets is one of the most investigated fields of database mining. The classic association mining based on a uniform support misses interesting patterns of low support or suffers from the bottleneck of itemset generation. A better solution is to exploit support constrains, witch specifies what minimum support is required for what itemsets and so that only necessary itemsets are generated. In this paper, an algorithm for multilevel database mining is proposed. The algorithm is obtained by extending the AFOPT algorithm for multi-level databases.
Keywords: data mining, knowledge discovery in databases, support constrains, multi-level databases, multi-level association rules mining

1 Introduction

The aim of data mining is the discovery of patterns within data stored in large databases. Mining for association rules is a data mining method that lends itself to formulating conditional statements such as "if customers buy product x, then they also buy product y". Market basket analysis is a typical example among various applications of association mining. The association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold [9].

Most studies on database mining have been focused at mining rules at single concept levels [2] and [9]. This approach may sometimes encounter difficulties of finding desired knowledge in databases. Frequent pattern mining plays an essential role in mining associations, as shows Agrawal and Srikant [2] in multi-level and multi-dimensional patterns, and many other important database mining tasks. Mining frequent patterns in transactions databases, multi-level databases, and many other kinds of databases has been studied popularly in data mining research. For mining multiple-level association rules, concept taxonomy should be provided for generalizing primitive level concepts to high level ones.

2 Multiple concept levels

The use of conceptual hierarchies facilitates the mining of multiple level rules. Moreover, conceptual hierarchies can be adjusted dynamically to meet the need of the current data mining task. Example: numerical attributes can be generated automatically based on the current data distribution. Compared to single level frequent pattern mining, multilevel frequent pattern mining is fine-tuned to generate interesting frequent patterns spanning at multiple concept levels in multi-level databases [3]. It services the business needs much better. Patterns generated with concept hierarchies included are called multi-level frequent patterns. The classic AFOPT algorithm (Ascending Frequency Ordered Prefix-Tree) [10] uses a compact data structure to represent the conditional databases, and the tree is traversed top-down. The combination of the top-down traversal strategy and ascending frequency order minimizes both the total number of conditional databases and the traversal cost of individual conditional databases. Based on the concept hierarchy, the AFOPT data structure [10] and some existing algorithms for mining single-level association rules [11], an efficient algorithm for mining multi-level frequent pattern is propose in this paper. We assume that the database contains: 1) An item data set which contains the description of each item in I in the form of Ai; description, where Ai is the product's code; 2) A transaction data set, T={T1,..., Tn}, which consists of a set of transactions Ti={Ap,...,Aq} in the format of (TID, Ai); Where TID is a transaction identifier and Ai is an item from the item data set [11]. Multi-level databases use hierarchy-information encoded transaction table instead of the original transaction table [2]. This is useful when we are interested in only a portion of the transaction database such as food, instead of all the items. This way we can first collect the relevant set of data and then work repeatedly on the task-relevant set.

Thus in the transaction table each item is encoded as a sequence of digits. Table 1, contains category codes and a description for each code (category or item) which is only needed for the final display. The codes and the descriptions for table 1 are extracted from figure 1.

TABLE 1.Multi-level items

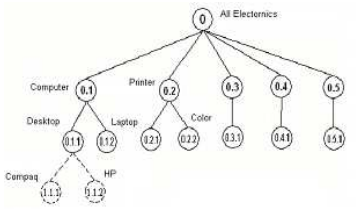


Figure 1: Hierarchy-base multi-level database

Code	Description
0	All Electronics
1	Computer
2	Printers
3	Scanners
4	Monitors
5	Keyboards
1.1	Desktop
1.2	Laptop
2.1	Compaq
2.2	HP
3.1	Mustek
4.1	Philips
5.1	Tech
1.1.1	Athlon
1.1.2	Pentium

The transactions from the database usually contain items from the lowest concept level. In order to get results for a certain concept level, we have to transform the transactions (or view them in a different way). This can be done by using the information from each item's ID.

TABLE 2 Transactions from database

TID	Items
T1	3.1.1 1.2.2 2.2.3 2.1.1 1.1.4 1.1.2
T2	2.1.2 4.1.1 1.2.2 1.1.2 3.1.1
T3	2.1.2 4.1.2 5.1.3 1.1.2
T4	1.2.3 1.1.2 3.1.1 4.1.1
T5	1.1.2 3.1.1 1.2.3 1.2.2

Multi-level frequent pattern mining is a very promising research topic and plays an invaluable role in real life applications. The classic frequent pattern mining algorithms (APRIORI [2], FP-Growth [8]) have been focusing on mining knowledge at single concept levels. It is often desirable to discover knowledge at multiple concept levels that are interesting and useful. A method for mining multiple-level association rules uses a hierarchy information encoded transaction table, instead of the original transaction table, in iterative data mining.

3 ADAPTIVE AFOPT algorithm - ADA AFOPT

In this paper, we propose an algorithm, named ADA AFOPT, which can be used to resolve the multilevel frequent pattern mining problem. The algorithm is obtained by extending the AFOPT algorithm for multi-level databases [8]. The features of the AFOPT algorithm, i.e. FPTree[6], FPTreebased pattern fragment, 1item partitionbased divide and conquer method are well preserved.

This algorithm uses flexible support constraints. To avoid the problem caused by uniform support threshold, mining with various support constraints is used. Uniform support threshold might cause problems of either generating uninteresting patterns at higher abstraction level or missing potential interesting patterns at lower abstraction are level. At each level, we classify individual items into two categories: normal items and exceptional items.

At each level, 2 types of thresholds, for both normal items and exceptional items, are needed. The two types of thresholds are "item passage threshold" and "item printing threshold". Usually item printing threshold is higher than item passage threshold. An item is treated as a frequent 1item if its occurrence passes the item printing threshold. Otherwise, as long as its support passes the item passage threshold, the item will still have les chance to be held in the base to generate frequent 2itemsets. This is reasoned on the observation that longer patterns are likely associated with smaller thresholds ion than their subpatterns.

ADA AFOPT algorithm favours users by pushing these various n transactions, support constraints deep into the mining process. The interestingness of the patterns mound generated, hence, is improved dramatically. Being based on the AFOPT algorithm, this algorithm first traverses the original database to find frequent items or abstract levels and sorts them in ascending frequency order. Then the original database is scanned the second time to construct an AFOPT structure to represent the conditional databases of the frequent items. The conditional database of an item i include all the transactions containing item i, and infrequent items and those items before i are removed from each transaction. Arrays are used to store single branches in the AFOPT structure to save space and construction cost. Each node in the AFOPT structure contains three pieces of information: an item id, the count of the itemset corresponding to the path from the root to the node, and the pointers pointing to the children of the node. In step 1 and step 2, the counts for each individual item is gathered; and the complete information about the transaction database is compressed in the AFOPT structure. From now on, the pattern generation process is started by recursively visiting the AFOPT structure. Note no more costly database its operations will be involved.

4 Performance study

The bottleneck of the Apriorilike algorithms [2], [4], [6], is the repeated database scans. The pattern growth approach avoids the candidate generation and test cost by growing a frequent itemset from its prefix [5]. Thus frequent pattern based algorithms are shown to be superior to the Apriori based ones significantly, especially on dense datasets. No matter how long the frequent pattern will be, it takes FPGrowth [8] a constant number of database scans (2 scans) to generate the problem complete set of frequent patterns. It is proved that FPGrowth is about an order of each magnitude faster than the Apriorilike algorithms. Using directly the AFOPT algorithm on the lowest concept level (or the level we need) will require only 2 scans of the database as opposed to k+1 scans with an Apriori based multi level algorithm. We traverse the AFOPT structure in top-down depth-first order. Each node is visited exactly once. The total number of node visits of the FP-Growth algorithm is equal to the total length of its branches. The total number of nodes visit of the AFOPT algorithm is equal to the size of the AFOPT structure, which is smaller than the total length of the nfp branches. Therefore the AFOPT algorithm needs less traversal cost than the FP-Growth algorithm For the AFOPT algorithm, additional traversal cost is caused by the push-right step. If we did not have this step we would get a conditional database which consists of multiple subtrees. The number of subtrees constituting the conditional database is exponential to the number of items before that item in worst case. While the number of merging operations needed is equal to the number of items before that item in worst case. To save the traversal cost, it is better to perform the merging operation. A set of experiments were conducted to compare the performance of the ADA AFOPT algorithm with the ADA FP-Growth algorithm. The standard versions of the FP-Growth and AFOPT algorithms were also part of the experiments. In the first test (fig.2), some items were considered rare special items, and special passage and printing thresholds were used. These thresholds are lower than the normal thresholds, because patterns containing these are interesting even though are less frequent. As a result the algorithm had to deal with more items and the "ADA AFOPT" and is slower than standard version which doesn't use special thresholds.

In the second test (fig.3), some items were considered "common" special items, and the "common" passage and printing thresholds were used on them. These thresholds are higher than the normal thresholds. Consequently the algorithm had to deal with less items and the "Adaptive AFOPT" performance is better than standard version which doesn't use such thresholds. No rare special items were declared

The same conditions were applied to the ADA-FP-Growth algorithm, and standard FP-Growth. As stated earlier, the AFOPT algorithm is faster than FP-Growth; as a result, ADA AFOPT is faster than the ADA-FP-Growth algorithm (fig.4). Note that, even though ADA AFOPT has more items to deal with, it is still faster than FP-Growth which uses standard thresholds

Experiments were conducted on a 1.6 Ghz Athlon XP with 512 MB memory running Microsoft Windows XP Professional, using a real database of 50000 entries in 6000 transactions. All codes were done in java and compiled using the Eclipse Platform.

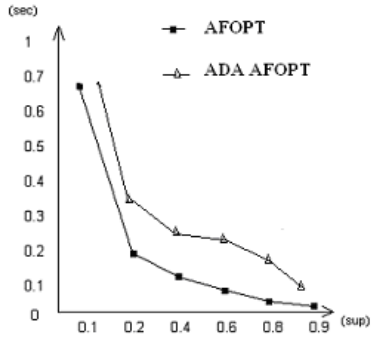


Figure 2: AFOPT vs. ADA AFOPT - Performance comparison with rare special items

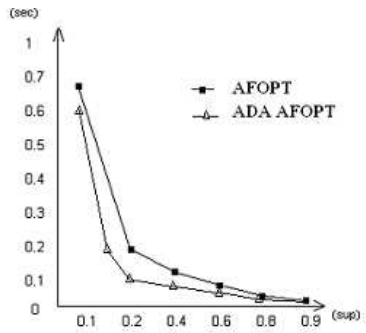


Figure 3: AFOPT vs. ADA AFOPT - Performance comparison without rare special items

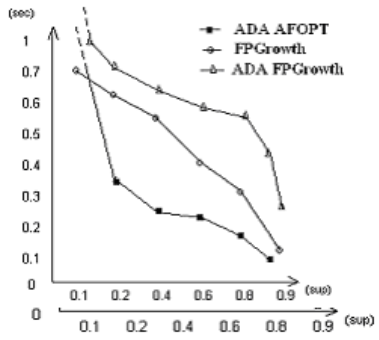


Figure 4: ADA AFOPT vs. FP-Growth vs. ADA FP-Growth - Performance comparison

5 Conclusions

Multilevel frequent pattern mining is fine-tuned to generate interesting frequent patterns spanning at multiple concept levels. It services the business needs much better and can be used to facilitate decision making and boost business sales. General frequent pattern mining algorithms focus on mining at single level. These way only strong associations between items will be discovered. For multilevel frequent pattern mining, mining algorithms have to be extended. The AFOPT algorithm traverses the trees in top-down depth-first order, and the items in the prefix-trees are sorted in ascending frequency order. The combination of these two methods is more efficient than the combination of the bottom-up traversal strategy and descending frequency order, which is adopted by the FP-Growth algorithm. The ADAAFOPT algorithm value lies on how to exploit these potential support requirements. The algorithm favours users by pushing these various support constraints deep into the mining process. The interestingness of the patterns generated, hence, is improved dramatically.

References

[1] Agarwal, R.C., Aggarwal, C.C., and Prasad, V.V.V., A tree projection algorithm for finding frequent itemsets, *Journal on Parallel and Distributed Computing*, 2001

[2] Agarwal R., Aggarwal C., and Prasad V.V.V., A tree projection algorithm for generation of frequent itemsets, In *J. Parallel and Distributed Computing*, 2000.

[3] Agrawal R., Mannila H., Srikant R., Toivonen H., and Verkamo A.I., Fast discovery of association rules, In *Advances in Knowledge Discovery and Data Mining*, pages 307-328, 1996

[4] Agrawal R. and Strikant R., Mining sequential patterns, In *Proc. 1995 Int. Conf. Data Engineering*, 3-14, Taipei, Taiwan, 1995

[5] Agrawal R. and Strikant R., Fast algorithms for mining association rules, In *Proc. 1994 Int. Conf. Very Large Data Bases (VLDB'95)*, Santiago, Chile, 487-499, 1994

[6] Bayardo R. J.., Efficiently mining long patterns from databases, In *SIGMOD'98*, pp. 85-93.

[7] Dong G. and Li J., Efficient mining of emerging patterns: Discovering trends and differences, In *KDD'99*, pp. 43-52.

[8] Gyorodi R., Gyorodi C., Pater M., Boc O., David Z., AFOPT Algorithm for multi-level databases, *SYNASC 05*, Timisoara, 2005

[9] Han, J., Pei, J., and Yin, Y. 2000, Mining frequent patterns without candidate generation, In *Proceedings of the 2000 ACM SIGMOD Conference*, ACM Press, pp. 1-12.

[10] Liu, G., Lu, H., Lou, W., Xu, Y. and Xu Yu, J., Efficient Mining of Frequent Patterns Using Ascending Frequency Ordered Prefix-Tree, *Data Mining and Knowledge Discovery*, 9, 249-274, 2004

[11] Liu, G., Lu, H., Lou, W., Xu, Y. and Xu Yu, J., Ascending Frequency Ordered Prefix-tree: Efficient Mining of Frequent Patterns In *Proc. of KDD Conf.*, 2003.

[12] Zaki, M.J., Parthasarathy, S., Ogihara, M., and Li,W. 1997, New algorithms for fast discovery of association rules, In *Proceedings of the Third KDD Conference*, AAAI Press, pp. 283-286.

[13] Zheng, Z., Kohavi, R., and Mason, L. 2001, Real world performance of association rule algorithms, In *Proceedings of the 7th KDD Conference*, ACM Press, pp. 401-406.

Mirela Pater, Daniela E. Popescu
University of Oradea
Department of Computer Science
University Street, no.1
E-mail: {mirelap,depopescu}@uoradea.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 442-447

Object-Oriented Construction of Portals Using AJAX

Alexandru Florin Pavel, Crenguţa Mădălina Bogdan

Abstract: In the matter of a very few years, portals have become complex and powerful web-based systems, which provide to the users both centralized access to information from different sources, and a great range of services like e-mail, searching, forum, and so on.
In this paper, we present a way to develop portals conforming to the principles of the object-oriented methodology and to construct 3-tier software architectures for portals. In order to do this, we use the MVC (Model-View-Controller) architectural pattern and depending on their responsibilities, we furthermore classify the model objects in four categories: domain, entity, utility, and manager objects. Furthermore, the layered software architecture is implemented using the AJAX programming technique and C# language.
Our solution is exemplified on the development of a portal of an IT company, which makes on-line advertisements of its products.
Keywords: portal, software architecture, MVC, AJAX

1 Introduction

Portals were born from the people's need of real-time and centralized access to information. By definition, a portal is a web-based application that provides an aggregation of content from different sources. Such application is constructed based on the idea of a single point of access for the users and provides them a great area of services like e-mail, searching, e-commerce, forum, and so on. As a web-based application, a portal can channel the users to other web content or services outside the domain of the portal application, giving to the user the impression that he or she accesses only the services of the portal. In order to do this, a portal application must communicate with e-mail servers, database servers or other processing environments.

Today, most portals are constructed on two levels: web interface and database, such that objects, which manage the client's requests, contain instructions that access the database and provide the requested data. As the main contribution of this paper, we propose the construction of portals such that the following objectives are fulfilled: a) applying the software engineering principles of the object-oriented methodology and the models of the software process [5] to the portal development in the same manner in which it is developed in any software system that provides a web graphical interface to the users through which they can interact with the system; b) the portal has 3-tier software architecture, i.e. it is layered on three levels: web interface, business logic, and database. In order to design 3-tier software architecture we used the Model-View-Controller (MVC) architectural pattern (Section 1.1); c) fulfilling of the non-functional requirements or quality attributes of the portals, like elimination of page post backs, parallelization of tasks, limited bandwidth and so on. Some of these quality attributes are fulfilled by using the AJAX technique (Section 1.2); d) constructing of the web interface of the portal as an aggregation of partial pages, where each page represents a graphical view of the portal in order to fulfill some functional requirement (Section 2); e) reusing code as much as possible (Section 2).

We solved the above objectives by developing the process of a B2C (business-to-consumer) portal of an IT company. Through the portal, the company wants to advertise on-line their products, to pre-marketing of the products with the objective of knowing how well the products are welcomed on the market, to realize a connection with the company clients and to provide information support for the use of their IT products. The results are presented in Section 2.

1.1 MVC Architectural Pattern

Software architecture is a description of the components that form the software system, their relations to each other that coordinate the actions of these components, and the principles guiding its design and evolution [6]. The components can be modules, objects, web services, and so on, depending on the partitioning criterion used. In this paper, we will construct a object-oriented software architecture.

In general, the software is designed applying an architectural pattern. This describes the kind of components, their relations, their constraints, the design and the composition rules of the components.

In our approach, we use the MVC architectural pattern [1]. This pattern classifies the objects in three categories: view, model and controller objects. The classification criterion is given by the responsibilities of the objects from

each category. View objects are objects with which user stakeholders interact directly, such as frames, forms, or any objects that help to the construction of graphical interface. The model objects eventually contain persistent information managed by the portal. Many such objects come from the business objects of the domain model of the information system where the software system will operate. However, other objects from this category can also emerge during the design activity of the software architecture. Finally, the controller objects have the responsibility to manage the logical flow and the events produced by users in their interactions with the view objects.

The rules that constrain the communications between objects are given in the following: a) the objects from the same level can communicate between each other; b) user stakeholders can access only the view objects; c) view objects can communicate only with the controller objects. However, there are cases when the view objects send messages to the model objects, but these messages query their states and do not modify them; d) controller objects can communicate with the view objects; e) model objects communicate only with the controller objects.

In order to obtain quality software architecture the designer must apply the design principles of low coupling, high cohesion, and assignment of responsibilities. These principles are fulfilled if we use the general responsibilities assignment patterns like Information Expert, Creator, Low Coupling, High Cohesion, and Controller [3]. Other design patterns can also be used [1].

As modeling language, we use the Unified Modeling Language (UML), which is a standard modeling language used in the analysis and design of the information and software systems [4]. In our approach, we used UML during the software analysis and design of the portal.

1.2 AJAX Programming Technique

AJAX or Asynchronous JavaScript and XML is a new programming style or technique that is based on existing standards and technologies for creating faster and more interactive web applications [2]. The core of AJAX is XMLHttpRequest that gives web applications the ability to exchange data asynchronously with the web server eliminating the need for a complete page refresh. Apart from the XmlHttpRequest api AJAX uses a combination of XHTML or HTML and CSS, JavaScript, DOM and XML which is the preferred markup language for data formatting. Because web pages are loosely coupled and they need to be formatted before they are presented to the user they have to be re-loaded each time a user makes a http request. Using AJAX the page re-loading can be minimized or can be eliminated, as we will use in our approach.

One of the main advantages of using AJAX is the minimization of bandwidth usage. The “transfer only what you need” was impossible to implement in classic web architectures because the data needed to be marshaled into a recognized format, such as html and the page needed to be re-loaded each time a request for a new dataset is made. Implement an AJAX based solution makes possible the asynchronous exchange of data between the client browser and the web server. The data transfer is done in the background and is formatted using JavaScript on the client side. This implementation also reduces the consumption of hardware resources on the server side because the formatting is done on the client’s machine.

2 Our Solution

Because of limited space, we present our solution and in parallel we use the IT products portal case study, that was presented in the introduction.

During the software analysis phase, we identify the software use cases after the services requested by the software actors (like user, client, administrator, technician, and moderator) and the domain objects used like account, product, product category, forum, feedback, news, support, card, and rss document. In the case of our case study, we obtained 36 software use cases, which describe the behavior of the portal and its interaction with users with the scope to provide the services requested by the latter. For example, the users, clients and subscribers can search and view information about the IT products managed by the portal. Then, these software actors communicate with the portal during the execution of the search and view products use case.

In the software design phase, the portal architecture is modularly constructed, one package for each software use case. After that, in order to construct the architecture we applied the MVC architectural pattern, design patterns, and the following main heuristics:

- each and every view object is associated with a controller object that manages the events produced by the user using the controls of the view object. The default page has also a controller called in our portal BaseController;

- the model objects are classified in four categories: a) domain objects, such as Product, ProductCategory, i.e. objects that contain persistent information; b) manager objects which deal with the database in order to execute operations on persistent information; c) entity objects, that is those objects which do not belong to the other three categories, but they are necessary for the portal operating; and d) utility objects that manages that functionalities which are related by the particularities of the portal. For instance, like any web-based application, the portal has a class (in our case, the SessionState utility class) that ensures the integrity of the data by avoiding the deleting or modifying the contained data. Another example of utility classes is the SingletonConnection class. This class applies the Singleton design pattern to provide a unique object of the SqlConnection class that creates a connection to the database. To ensure that a unique object is created, we used another object (referred by the syncRoot variable) to obtain a critical section in which we create the object of the SingletonConnection class.

For instance, the design class diagram of the package that corresponds to the Search and view products use case is presented in Figure 1.

In conclusion, in order to obtain 3-tier portal software architecture, we classify the model objects in four categories. The first category is constituted by the domain objects that contain persistent information. The second one is formed by the utility objects that deal with some functionality related by the particularities of the portal. From the third category we have the manager objects, which contain operations that update or query the database. The last category contains that objects which are necessary for the portal functioning, but do not appertain to other categories. Such an object (called an entity object) is not associated with an object manager to store the information contained in the database or to interact with the information read in the database. Some objects from these categories together with the view and controller objects of the IT products portal are presented in Figure 2.

Furthermore, we divide the architecture objects in two subsystems: client and server. Conforming to the principles: some of the view objects together with the associated controller objects appertain to the client subsystem and other ones are used by the server subsystem; the manager objects and all the domain, entity or utility objects used by them stay on the server.

In order to implement the portal architecture, we used AJAX programming technique and the C# programming language. The view objects have been implemented as partial classes in C# and referred in HTML files. These files form partial pages, which are displayed in the portal web interface. Thus, the portal interface becomes an aggregation of partial pages displayed in some well-defined order given by the objects behavior.

The controller objects from the client subsystem are direct or not subclasses of the UserControl class and are used by the AJAX engine to communicate with the objects from the server. The other objects categories have been coded as instances of classes in C#. In Figure 3 we present the web interface of the IT products portal.

3 Conclusions

In this paper, we presented a way to construct a 3-tier software architecture of portals. The idea of the creation of a such solution has come from the challenge to object-oriented develop a portal like any software system that provides a web graphical interface to the users through which they can interact with the system. Although, we must take into account the particularities of the portals.

The second contribution is that we implemented the portal software architecture using the AJAX programming technique. This technique allows us to create the partial pages of the portal and to reduce the communication traffic between the client and server.

Acknowledgements

This work is funded within the TOMIS project, no: 11-041/2007, by the National Centre of Programs Management, PNCDI-2 - Partnerships program.

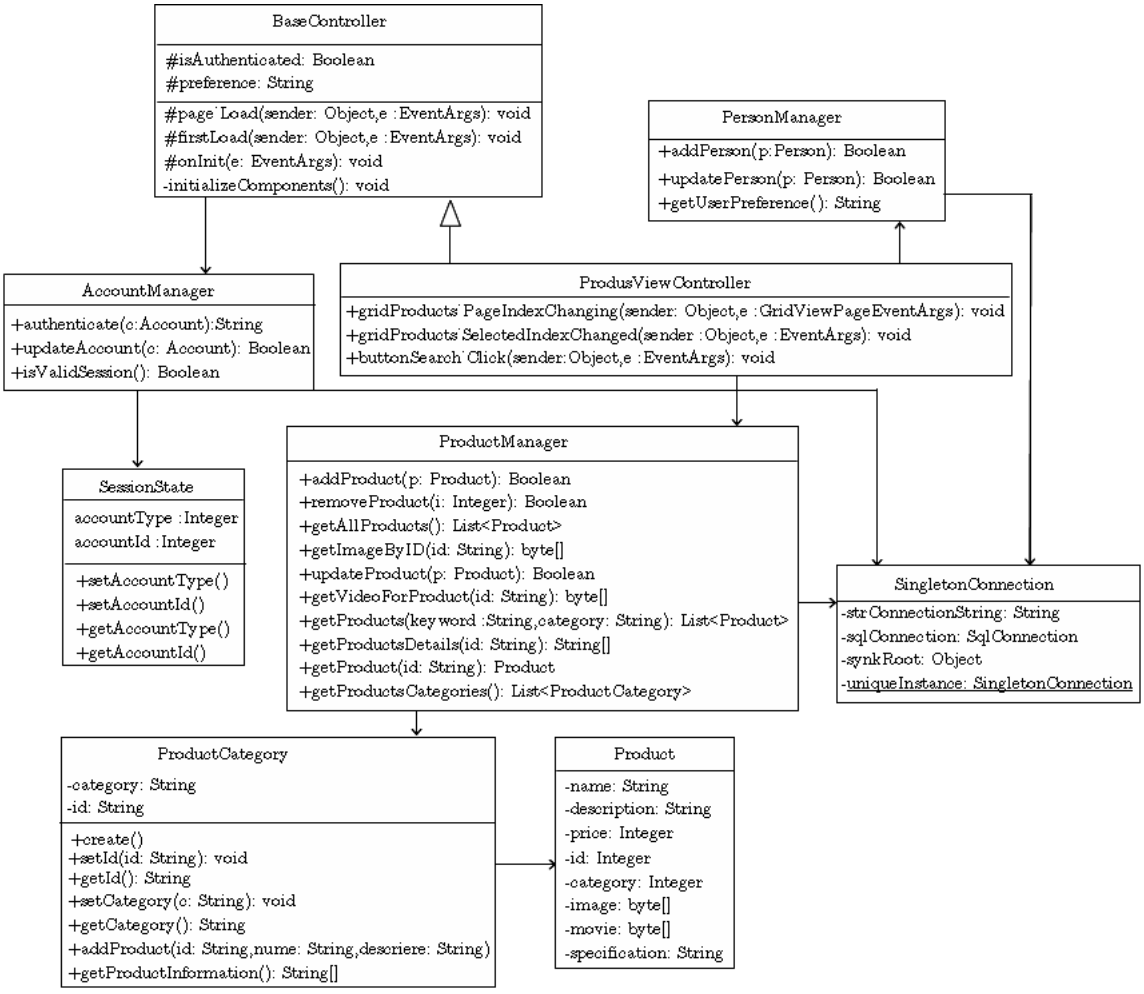


Figure 1: The class diagram of the Search and view package

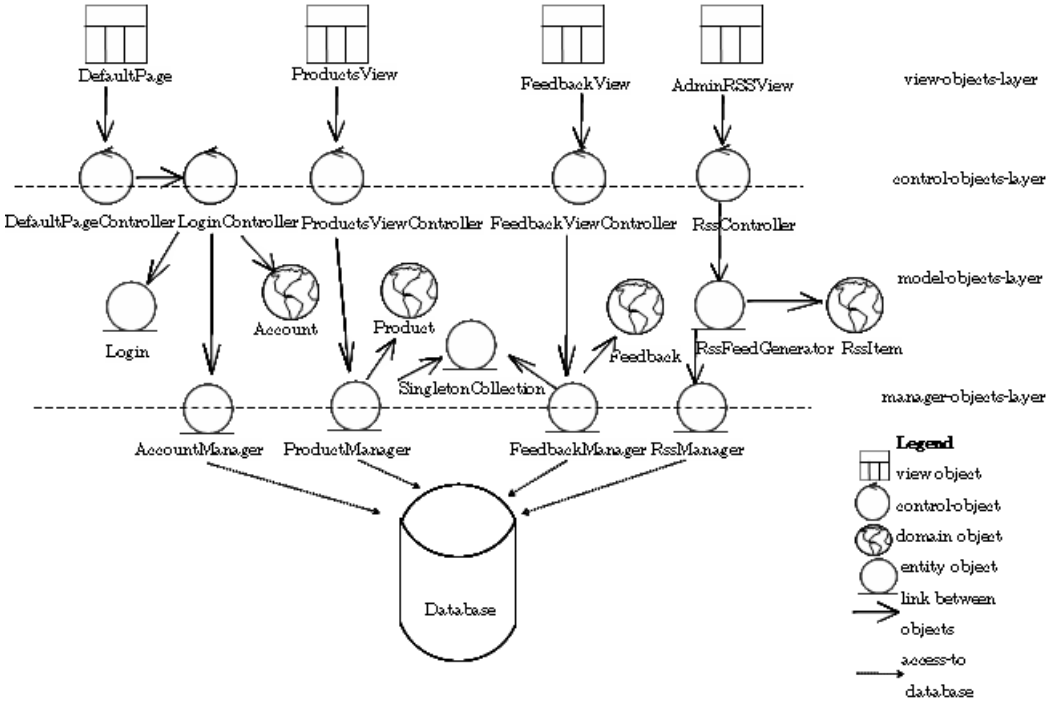


Figure 2: A part of the IT products portal software architecture

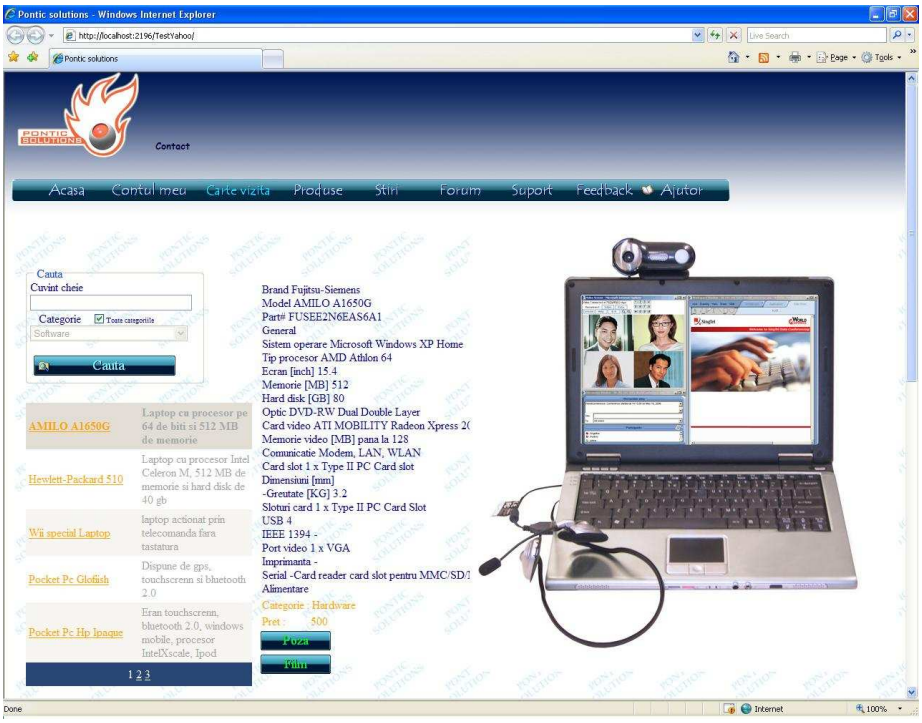


Figure 3: The interface of the IT products portal

References

[1] E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison Wesley Professional, 1994.

[2] M. Hertel, Aspects of AJAX, published online at http://www.mathertel.de/AJAX/AJAXeBook.aspx, 2007.

[3] C. Larman, *Applying UML and Patterns. An Introduction to Object-Oriented Analysis and Design and the Unified Process*, Prentice Hall, 2004.

[4] OMG, Unified Modeling Language Superstructure, version 2.0, ptc/03-0802, 2003.

[5] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, New York: McGraw-Hill, 2000.

[6] The Institute of Electrical and Electronics Engineers Standards Board, Recommended Practice for Architectural Description of Software-Intensive Systems (IEEE-Std-1471-2000), 2000.

Alexandru Florin Pavel
University of Sydney, Australia
Faculty of Engineering & Information Technologies
E-mail: alexandruyo@yahoo.com

Crenguţa Mădălina Bogdan
Ovidius University, Romania
Numerical Methods and Computer Science Department
124 Mamaia Blvd.
E-mail: cbogdan@univ-ovidius.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 448-453

A Distributed Simulation Framework for Mission Critical Systems in Nuclear Engineering and Radiological Protection

A. Piater, T.B. Ionescu, W. Scheuermann

Abstract: An analysis of the current state of mission critical simulation software in the domain of nuclear sciences is presented. The drawbacks of the KFÜ-ABR and VIP systems are pointed out and a new generation distributed simulation framework (DiSiF) is proposed. The modern, aspect oriented design of the new framework relies on the resource paradigm, actor based workflow modelling, and Web services and Grid computing, as implementation technology. DiSiF will eliminate the lack of scalability, flexibility, fault tolerance, and other disadvantages of the KFÜ-ABR system, currently in use by the Ministry of Environment of the German State of Baden-Württemberg.
Keywords: simulation, nuclear sciences, distributed systems, aspect oriented programming, actor based modelling.

1 Introduction

In the domain of nuclear sciences, simulation systems are used to give answers to many different physical aspects. Over the years many simulation programs have been developed implementing existing knowledge of these physical processes and their dynamic behavior. Due to the age of these programs they are not easy to use and it is difficult to run them on modern IT structures. To make these programs reusable and to preserve the contained knowledge, a new system architecture based on modern IT structures has to be developed.

In Germany the regulatory authorities use software systems for the 24/7 distance observation of nuclear power plants. In the German State of Baden-Württemberg this system is called "Kernreaktorfernüberwachung" (KFÜ). Besides measuring devices and database servers the KFÜ also contains a simulation system called "Ausbreitungsrechnung" (ABR) which calculates the release, the airborne transportation, and deposition of radioactive nuclides. The system is intended to be used in case of an accident to assist decision makers as they take the necessary steps for the protection of people and environment. Simulation systems used for such emergencies are mission critical. The current system fulfills the basic requirements of stability and ease of use in case of emergency but is limited to a single server instance. This also limits the system's scalability to a disproportionately expensive multi processor server. The final possible extension, without major improvements of the existing system architecture, is to switch from a 32 bit to a 64 bit environment.

In the context of this paper a new architecture will be introduced that offers methods to integrate legacy simulation modules in distributed and mission critical simulation systems. Based on this new architecture a framework is implemented on which the next generation of the simulation system as part of the KFÜ in Baden-Württemberg will be based. It will offer major improvements in the way existing simulation modules can be integrated into the simulation framework, resulting in: a higher level of fault tolerance; a better scalability; a better integration of legacy codes, i.e. existing simulation modules; easier extension of the simulation modules for non-programmers; re-use of the simulation system in other contexts; re-use of the framework to build simulation systems in other engineering disciplines.

To overcome the bottleneck of a single server instance and to introduce the possibility of a scaling out, this framework will provide a distribution mechanism which will improve fault tolerance if a server breaks down. A general approach for such a new framework is introduced which is not limited to the ABR simulation system. It can also be used for the improvement of similar simulation systems which can be found in many engineering disciplines and sciences. Therefore a second example is given which shows the usability of this framework in the Virtual Power Plant (VIP) [2] simulation system. Chapter 2 works out the similarities between these two legacy simulation systems. This leads on to chapter 3 where details of the new Distributed Simulation Framework (DiSiF) are shown. It describes the framework itself and introduces the resource paradigm on which layered resources can exchange e.g. session, simulation, and workflow information. Furthermore it describes the use of a Aspect Oriented Design which is used for security reasons and also covers the use of Kepler [4], Condor Grid software, and web services for being technology neutral.

2 Simulation Systems in the Domain of Nuclear Sciences

Figure 1 depicts the KFÜ system [1] which has been fully operational since the year 2000. The central position in the KFÜ system is taken by the Central Data Storage (CDS), a big database system which gets about 100,000 values per day, collected and delivered by the Communication Server (CS) from several measuring devices located directly in and around nuclear power plants. The Application Server (AS) is responsible for alerting the regulatory authorities in case of an emergency. It also triggers the ABR simulation system automatically. When the ABR is triggered it calculates the release, the airborne transportation, and deposition of radioactive nuclides. In case of an accident these calculations are mission critical. The ABR receives the measured data directly from the CDS via a web service interface [3]. The graphical user interface of the KFÜ system is the Client which connects via the Internet to the CDS. It is also possible for an user to employ the ABR system to perform simulation calculations manually. Therefore the Client can also directly connect to the ABR simulation system.

The ABR system is a modular system which offers different simulation models depending on the required accuracy of the results and computing time. The existing simulation modules were written in FORTRAN, C, and C++ programming languages. The design of these simulation modules lacks in the way they can be integrated in distributed simulation systems. This is because the modules can only understand proprietary data formats in form of text files.

ABR is limited to a single server instance which restricts the scalability to the number of CPUs and memory of such a machine. This also limits the possible complexity level of calculations. The simulation system is fault tolerant concerning missing measuring values but there is no redundancy if this single simulation server breaks down.

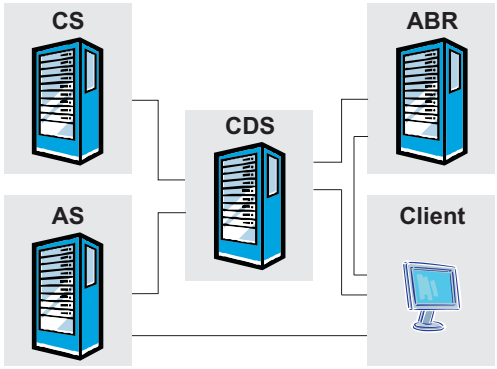


Figure 1: KFÜ system architecture overview. *CDS* = Central Data Storage, *CS* = Communication Server, *AS* = Application Server, *ABR* = ABR Simulation System, *Client* = KFÜ User Interface

To analyze the usability of the framework introduced in chapter 3 in other applications areas, a second legacy simulation system was investigated. The Virtual Power Plant (VIP), a simulation system used to calculate the transients of a high temperature reactor (HTR) core also consists of several simulation modules. These modules have been developed using FORTRAN and are only optimized on their physical behavior and their usage of system resources. Due to the fact that these modules were developed in the 1980's, the user interface is realized in a cryptic way using text files in a proprietary legacy data format. VIP has no designated workflow engine. The workflows are programmed manually and utilize batch processing. VIP offers no external runtime control abilities and is limited to a single machine instance during runtime which often results in long calculation times. There is no graphical user interface available and no access control is implemented yet. All visualizations are done manually based on the output files.

Table 1 shows the similarities and the differences between the legacy simulation systems ABR and VIP. It shows that these two systems have many points in common even if VIP doesn't make use of external data sources and mission critical runtime abilities. This means that a framework which is able to cover the needs of ABR will also be able to cover the needs of VIP.

3 A New Generation of Distributed Simulation Frameworks

The new generation of simulation frameworks in the domain of nuclear engineering must solve all of the issues mentioned above. Some of these issues were caused by the lack of suitable technology at the time these systems were developed, or the limitations imposed by legacy software. These new generation simulation systems will be based on flexible, scalable and easily maintainable software systems. The underlying architecture will provide adaptability to new domain-specific problems, like the one of extending the geographical area of the ABR system. In order to achieve these goals, the new system was conceived by first modeling the simulation problems on an abstract level and then by integrating the resulted scientific workflows into a distributed environment.

The Resource Paradigm. The classic simulation scenario in nuclear engineering is the one where a user has to perform a simulation in order to obtain certain results that can be used for further work. The entire simulation

	ABR	VIP
Legacy issues	✓	✓
Cryptic I/O file formats	✓	✓
Workflow dependency between modules	✓	✓
Only programmers can change workflows	✓	✓
Limited scalability	✓	✓
External data sources	✓	X
GUI	✓	X
Fault tolerant	O	X
Secured access control	✓	X
Mission critical runtime abilities	✓	X

Table 1: Overlappings and differences between ABR and VIP (✓= Yes, O = Partial, X = No)

is used as a simulation tool or resource by this user, who, most of the time, is unaware of the inherent algorithms and technicalities that make the simulation possible. The user employs the simulation resource and must make a decision based solely on the result of the complete underlying chain of operations. The simulation is composed of one or several workflows which play the role of algorithmic resources for the simulation. Finally, every workflow is composed of computational operations, that can be executed in parallel mode or sequentially and finally provide a result. In DiSiF these computational operations are regarded as the workflow's resources. We called this *a hierarchy of autonomous resources* and represented it as a layered pyramid (see figure 2). The relations between hierarchical autonomous resources have the following properties: *i.* Resources can exchange information with their neighbors only (i.e. upper, lower, same level layer); *ii.* A resources from layer*n* is the owner of the resources it has instantiated at layer*n-1*; *iii.* Information can only be exchanged between a resource and its owner or a resource and its owned resources.

Resource providers can be different software tools, databases or powerful computers, depending on the resource types. A certain type of resource can only run on certain types of servers which are suitable providers for that type of resource (e.g. a workflow resource needs to run on a machine with a powerful workflow execution engine installed). The following relations exist between resources and providers: *i.* A resource can have more than one provider of different types; *ii.* There can be more than one provider for a certain type of resource that is very often needed; *iii.* Several types of resources can have a single provider.

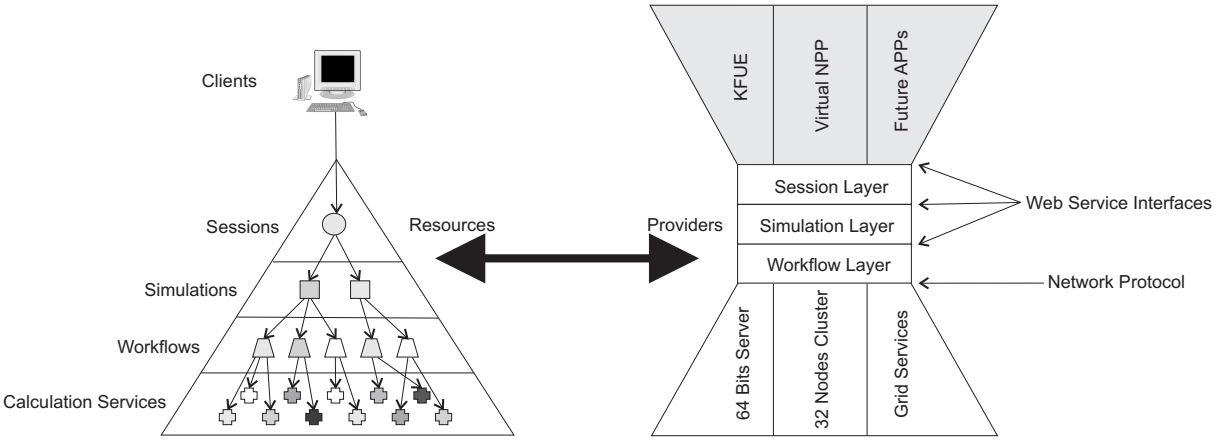


Figure 2: The architecture of DiSiF

Figure 2 also shows the layered hourglass software architecture supporting the abstract resource oriented model. Layered software models have been proven to be very efficient in the fields of networking and communication but also in almost all other software development fields, including embedded software [5] or the n-tier model. Each layer corresponds to a resource provider and can run on different machines. Objects that are instantiated at different levels in the stack are the actual resources. The waist of the hourglass is composed of three layers:

- *The Session layer* - Host of the client session resource; at this level the user role policy is applied and the

corresponding simulation resources are advertised and controlled;

- *The Simulation layer* - Host of the simulation resource which, in case of the ABR system, corresponds to a complete end to end propagation calculation; a simulation resource manages the execution of all the underlying scientific workflows;

- *The Workflow layer* - Host of the workflow resource that manages the execution of the underlying operational modules.

The fat top of the hourglass is represented by the different remote clients using the simulation framework through a thin adaptation layer, if necessary. The fat bottom is represented by the different employable job execution technologies that for the different computation steps of the scientific workflows. This flexibility in choosing the job execution environment is enabled by actor-oriented modelling [6].

An Aspect Orieted Design. Besides the need for the simulation to be distributed, another aspect that is extremely important, especially in the field of nuclear engineering, is security. User roles and execution rights are assigned on a per resource basis as well: the owner of a resource is the only one that can access and use the resource, the user being the ultimate (human) resource to which a role is assigned. The user interface can be automatically generated based on the resource access policy and the parameter dependency rules that govern the relations between different workflows of a simulation [8]. Another concern with distributed systems is the one of synchronicity. Each resource can be instantiated and its methods executed at different levels of the software stack, both in synchronous and asynchronous modes. Synchronous mode means that each time resource A calls a method of resource B, owned by A, the caller waits until the execution of the method finished. Asynchronous mode means that resource A calls a method of resource B, which immediately starts a new thread where the requested operation is executed and returns from the method. A notification is sent to A when the operation is finished.

So, there are basically three aspects that have to be taken into account - distribution, security and synchronicity - that don't directly concern the developers of scientific workflows or simulation designers. The creators of such models aren't usually programmers nor computer science engineers, but rather physicists or nuclear engineers, with little or no programming abilities. Therefore DiSiF uses an aspect oriented design [7] that supports the concept of separation of concerns [9], where all three aspects mentioned above are handled by the dispatcher component. The dispatcher is called every time a method of a resource object is invoked. It then applies the security policy, searches for a remote resource provider that can host the requested resource and then sends the request to the server. If an asynchronous method call is requested, the dispatcher manages the response by automatically triggering an event on the caller side, when the notification from the remote resource arrives.

Implementation details. Simulation software in nuclear engineering is always very heterogeneous from the technological and conceptual points of view and poses legacy problems. It is not uncommon for a single simulation system to encompass technology ranging from over 30 years ago to present times. One way to solve these problems and to make the system flexible and maintainable is to use web services. Web services also provide the prerequisites of the required distributed, security and synchronicity features for DiSiF. Moreover, using web services makes it possible for each layer of the system to be implemented using a different technology. For instance, one can use Kepler, which is Java based software, for the Workflow layer, the .NET framework for the upper layers (simulation and session) and the Condor Grid software for executing the low level computational jobs.

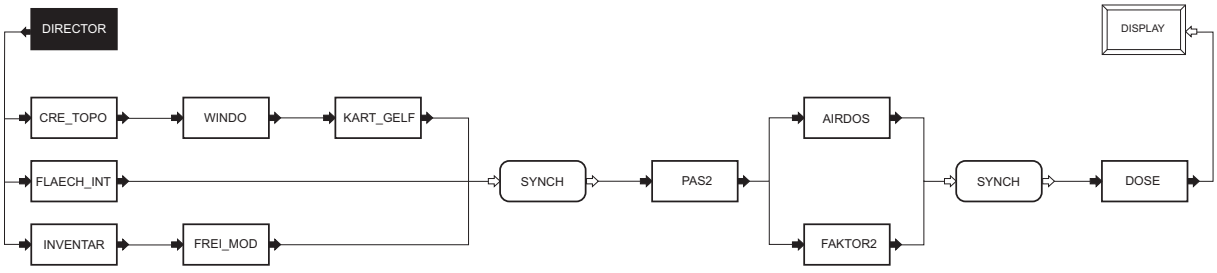


Figure 3: The ABR system's scientific workflow

Figure 3 shows one of the workflows of the ABR system which has been designed using Kepler. Each of the computational modules is represented by an actor. The model controls the information flow and synchronizes

the firing of the different actors, thus allowing automatic parallel or sequential execution of different branches of the workflow. The SDF (Synchronous Data Flow) director fires the open end actors, i.e. it creates the required number of tokens in order to execute the workflow. The synchronizer actor waits for all the brunches to complete (consumes several tokens and forwards only one) before it fires the next actor(s).

The aspect oriented component of DiSiF, the dispatcher, can be implemented in any programming language that supports web services and reflection [10]. Whereas *AspectJ* would provide a solution for a Java implementation of the dispatcher, it is also possible to implement aspects without using an AOP add on by simply adding dispatcher code before the actual implementation of each method of the resource classes, an operation that could also be done automatically by a customized code generator.

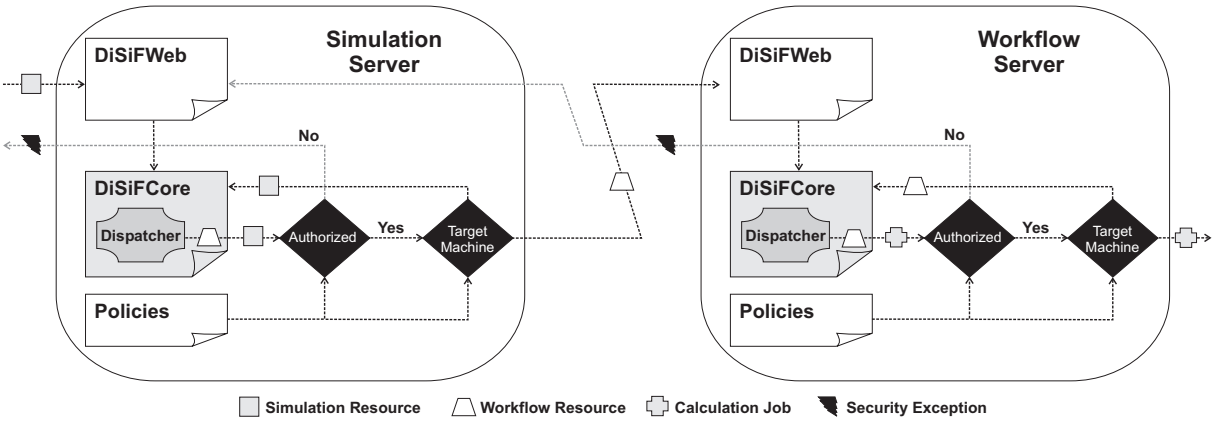


Figure 4: The dispatching mechanism used by DiSiF

Figure 4 shows the dispatching mechanism that allows the use of distributed resources. The dispatcher assesses, based on the local security policy, whether the call is permitted or not. If it is permitted, the distribution policy is applied and the call is forwarded to a resource on the specified server. If the *TargetMachine* happens to be the local machine, then the code is actually executed, else the request is forwarded to another machine. The entire functionality of the system is contained within a single library, *DiSiFCore*, that is distributed over a simple web service, *DiSiFWeb*, using reflection.

The presented technique has several advantages: it prevents inexperienced programmers from deviating from the resource paradigm when implementing new features and resources; it uses standard Web Service technology only; it allows for different methods of a class to be executed on different machines by transferring the entire binary serialized object from one machine to another.

4 Summary and Conclusions

A new generation distributed simulation framework (DiSiF) for the nuclear sciences domain has been presented. DiSiF makes using existing simulation programs easier and thus preserves the knowledge embedded in these programs as well as downsizing the learning curve in using this programs. It also offers the opportunity to use simulation systems in other contexts, like academic or training courses.

With its clear structure, the flexibility and availability will be increased in many areas, e.g. distribution over different calculation nodes, in the combination or exchange of modules within simulation models as well as the ability to offer well adapted user interfaces for each context the simulation systems will be used in.

The use of the Kepler allows non-programmers to easily model and design workflows without needing to know any programming languages, whereas web services eliminate most of the technological overhead that plagued older systems, like ABR and VIP.

References

[1] M. Weigele, F. Schmidt, K. De Marco, Ch. Krass, D. Sucic, D. Wagner, R. Chaker, R. Obrecht, G. Kaufhold, K. Zetzmann, R. Micheler: ABR-KFÜ - Der Dienst Ausbreitungsrechnung in der Kernreaktor-Fernüberwachung Baden-Württemberg.

In: *UIS Baden-Württemberg. R. Mayer-Föll, A. Keitel, W. Geiger (Hrsg.), Projekt AJA Phase II 2001, Forschungszentrum Karlsruhe GmbH, Karlsruhe*, pp.115-132, ISSN 0947-8620, 2001.

[2] M. Weigele, J. Achenbach, A. Piater, F. Schmidt, A. Schulz, D. Sucic, R. Obrecht, S. Weimer, R. Bechtler: KFÜ-ABR Entwicklung eines ABR-Research Systems. In: *UIS Baden-Württemberg. R. Mayer-Föll, A. Keitel, W. Geiger (Hrsg.), Projekt AJA Phase IV 2003, Forschungszentrum Karlsruhe GmbH, Karlsruhe*, pp.111-132, ISSN 0947-8620, 2003.

[3] A. Piater, W. Scheuermann, C. Krass, D. Wagner, R. Obrecht, H. Pohl: ABR-Research KFÜ - Anbindung an die zentrale Datenhaltung der Kernreaktor-Fernüberwachung Baden-Württemberg zur Durchführung von Prognoserechnungen. In: *UIS Baden-Württemberg. R. Mayer-Föll, A. Keitel, W. Geiger (Hrsg.), F+E-Vorhaben KEWA Phase II 2006/07, Forschungszentrum Karlsruhe GmbH, Karlsruhe*, pp.135-142, ISSN 0947-8620, 2007.

[4] I. Altintas, C. Berkley, E. Jaeger, M. Jones, B. Ludscher, S. Mock: Kepler: An Extensible System for Design and Execution of Scientific Workflows. *16th Intl. Conference on Scientific and Statistical Database Management (SSDBM)*, Santorini Island, Greece, June 2004.

[5] H. Heinecke, et al.: AUTomotive Open System ARchitecture - An industry-wide initiative to manage the complexity of emerging Automotive E/E-Architectures Convergence 2004. *International Congress on Transportation Electronics*, Detroit, 2004.

[6] S. Bowers, B. Ludäscher: Actor-Oriented Design of Scientific Workflows. *24th Intl. Conference on Conceptual Modeling (ER'05)*, Klagenfurt, Austria, LNCS, Springer, 2005.

[7] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. Lopes, J.M. Loingtier, J. Irwin: Aspect-Oriented Programming. *Proceedings of the European Conference on Object-Oriented Programming*, vol.1241, pp.220-242, 1997.

[8] A. Piater: Entwicklung eines Rollenmodells zur nachhaltigen Unterstützung der Forschung und Lehre im Bereich Kerntechnik. *Diss.*, Stuttgart, ISSN-0173-6892, 2007.

[9] E. Dijkstra: On the role of scientific thought. In *Dijkstra, Edsger W., Selected writings on Computing: A Personal Perspective*, New York, NY, USA: Springer-Verlag New York, Inc., pp. 60-66, ISBN 0-387-90652-5, 1982.

[10] J. Sobel: An Introduction to Reflection-Oriented Programming. *Reflection '96*, 1996.

A. Piater, T.B. Ionescu, W. Scheuermann
University of Stuttgart
Institute of Nuclear Technology and Energy Systems (IKE)
Pfaffenwaldring 31, D-70569 Stuttgart, Germany
E-mail: {piater,tudor.ionescu,scheuermann}@ike.uni-stuttgart.de

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 454-459

A Performance Comparative Analysis for Three Different Topological Tests Generation Algorithms

Daniela E. Popescu, Mirela Pater

Abstract: After describing the basic concepts of the three topological algorithms: D, PO-DEM and FAN, which we were use for developing the test generation modules of our software testing tool (TDL), the paper presents some comparative results concerning the adders testing - obtained by using this tool together with our conclusions regarding the advantages of every considered algorithm.
Keywords: test generation algorithm, faults, testing

1 Test pattern generation for structured algorithms

Before presenting the algorithms considered, we will explain some common concepts bound to the subject.

A singular cube is called each prime implicant of the singular cover. A primitive D-cube of a fault is used to express tests for a fault in terms of the input and output lines of the faulty gate. As an example, consider the AND gate with an output s-a-0 fault shown below. To test the fault it is necessary to set 1 on the output line, and thus lines 1 and 2 must be forced to value 1. Then line 3 will have value 0 if the fault is present and value 1 if it is absent. Hence, the corresponding primitive D-cube of the fault is given in Figure 1:

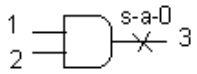


Figure 1:

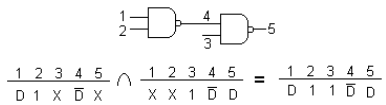


Figure 2:

where the value D means value 1 in the normal circuit and value 0 in the faulty circuit. Similarly, will be used to represent a signal that has the value 0 in the normal circuit and 1 in the faulty circuit. And the third type of cube is the D-intersection, which is used to generate a sensitized path. To explain this cube we present in Figure 2 an example in which we want to generate sensitized path from line 1 to line 5 and we will D-intersect the relevant propagation D-cubes for the two NAND gates [3].

2 The D Algorithm

The purpose of any test stimuli vectors generation algorithm is to find a test vector for a certain specified fault, which vector is applied on primary entries circuit to mark out the fault.

The single-path sensitization has a fatal flaw because there can exist testable faults in a circuit for which it is impossible to generate tests if only one path is allowed to be sensitizated at a time. Therefore, in order to guarantee finding a test for a fault if one exists, we must consider the sensitization of all possible paths from the site of the fault to the circuit outputs simultaneously. To do this, the D algorithm is formalized in terms of a cubical algebra called the D-calculus. A characteristic feature of the D algorithm is its ability to propagate errors on several reconvergent paths and it referred to as multiple-path sensitization [1].

3 PODEM Algorithm

Because the D algorithm was proved less efficient to solve the requires for those circuits named "ECAT" a class of combinational circuits that implements the functions of "correction - error - and - translation" - characterized by

a great re-convergence, Goel et. al. proposed a new algorithm named PODEM, (Path Oriented DEcision Making).

Differing of D algorithm, PODEM try to reduce the returning numbers toward the same decisions (so-called backtrackings). Those backtrackings are more numerous in case of D algorithm application because this allow attribution of values to internal lines and so returning process, to anterior decision can appear to the level of each gate; PODEM algorithm allow the attribution of value only for primary entries, and those values will be than conducted toward the circuit exit through the implication process [2]. PODEM algorithm, implicit tests the all patterns for main entries, but exhaustive those groups which could build the set of the test stimuli vectors. Also is built another decision graph whose knots are points on the way which the algorithm try to establish toward primary exists for propagate the fault. For sure, the primary entries are ending as soon as is found a test stimuli vector. If nor one pattern of the primary entries could give a test vector, it is considered that the fault is undetectable or redundant. Figure 7 gives the PODEM algorithm.

Although searched test vector should belong the space of all the vectors that can be applied on the primary entries, test stimuli vectors searched by the D algorithm belongs to another space. In case of the D algorithm the decision consist from the selection either of one gate from the D front or one way to justify the value applied to the entry of one gate belonging to the J front. Those decisions will determinate in the end the value of the primary entries, but properly searching is indirectly one.

The PODEM algorithm is an algorithm characterized by a direct search process that is trying a fast way search to the value which we must to consider to the primary entries for mark out the fault.

PODEM algorithm handles a logic value vk as being justified for k line with the objective (k, vk) that must be realized toward the primary entries. [6] gives the "backtracking" procedure needed in the implementation of this algorithm:

```
Objective ( )
begin
    /* the target fault is I s-a-v */
        if (the value of I is X) then return (I, n(v))
        select a gate (G) from the D-frontier
        select an input (j) of G with the value X
        c = controlling value of G
        return (j, n(c))
end.
```

Figure 3:

where n (v) and n(c) represents the negation values of the v and c.

Figure 6 presents the procedure used by our PODEM algorithm, to select an objective, respectively a gate for whom to justify the lines which are intersected with that gate. Having those both procedure presented above, which define the PODEM algorithm analyze manner, Figure 7 presents the procedure which implements the PODEM algorithm.

```
PODEM ( )
begin
    if (the error has propagate to an primary output) then return
SUCCESS
        if (test can't be generate) then return FAILURE
        (k, v_k) = Objective ()
    (j, v_j) = Backtrace (k, v_k)    /* j is a primary input */
        Imply (j, v_j)
    if PODEM () = SUCCESS then return SUCCESS
    /* pick the negative value for the decision*/
    Imply (j, n (v_j))
        if PODEM () = SUCCESS then return SUCCESS
        Imply (j, X)
        return FAILURE
end
```

Figure 4:

4 The FAN algorithm

The PODEM algorithm organize a decisions graph which has one or more decisions for every knout. Initial decision is choosing arbitrary, but for sure at some moment will have to return to an anterior knout where from result the backtracking process. In order to reduce the test generation necessary time it was tried reducing the periods between those backtrackings. The proposed solution was that "look ahead" to insure that it is in vain to continue on a specific way if we know for sure that we won't find solutions. So, the FAN algorithm filled in two main modifications comparative to the PODEM algorithm: 1.It stops the backtracking at the level of the intern lines and doesn't go on until the primary entries are reached; 2.It tries to achieve more different objectives simultaneous. vFor this reason, the FAN algorithm uses a multiple-backtrace procedure, which purpose is to execute a set of objectives. Figure 6 presents the MultipleBacktrace procedure and Figure 7 gives the FAN algorithm:

```
MultiBacktrace (current_objectives)
start
       repeat
            start
     remove one entry (k, v_k) from the current_objectives
              if k is a head line
                     then add (k, v_k) to head_objectives
              else if k is a fanout branch then
         start
           j = stem (k)
                                      increment number of request at j for v_k
                                      add j to stem_objectives
                           stop
                     else
                           start
                                      i = inversion of k
           c = controlling value of k
           if (v =i * c) then
                                                start
               select an input j of k with value X
               add (j, c) to current_objectives
                                                stop
                                      else  for every input j of k with value X
                                                add (j, c) to current_objectives
                           stop
            stop
       until current_objectives=0
       if stem_objectives != 0 then
     start
       remove the highest stem k from the stem_objectives
                     v_k = the most request value for k
                     if (k has contradictory values and
                                    k is not reacheable from target fault)
                     then return (k, v_k)
                     add (k, v_k) to the current_objectives
                     return MultiBacktrace (current_objectives)
     stop
   remove (k, v_k) from head_objectives
       return (k, v_k)
       stop.
```

Figure 5:

5 The TDL Testing Tool

In order to realize the comparative analyze concerning the performance we have implemented a testing software named Testing Digital Language (TDL). The goal of implementing such a software environment was the development of a new CAD tool with strong testing facilities, which can be used in the designing process for IC with high testability facilities.

```
FAN ()
start
  if Imply_and_check () = FAILURE then return FAILURE
        if (error at PO and all bound lines are justified) then
          start
            justify all unjustified head lines
                  return SUCCESS
          stop
if (error not on PO and D_frontier = 0) then return FAILURE
        /* initialize the objectives */
        add every unjustified bound line to current_objectives
        select one of the gates G with value X
        c = controlling value of G
        for every input j of the G with value X
        add (j, not(c)) on current_objectives
/* MultiBacktrace */
        (i, v_i) = MultiBacktrace(current_objectives)
Asign (i, v_i)
if FAN () = SUCCESS then return SUCCESS
Asign (i, not(v_i)) /* reverse decision */
if FAN () = SUCCESS then return SUCCESS
Asign (i, X)
        return FAILURE
stop.
```

Figure 6:

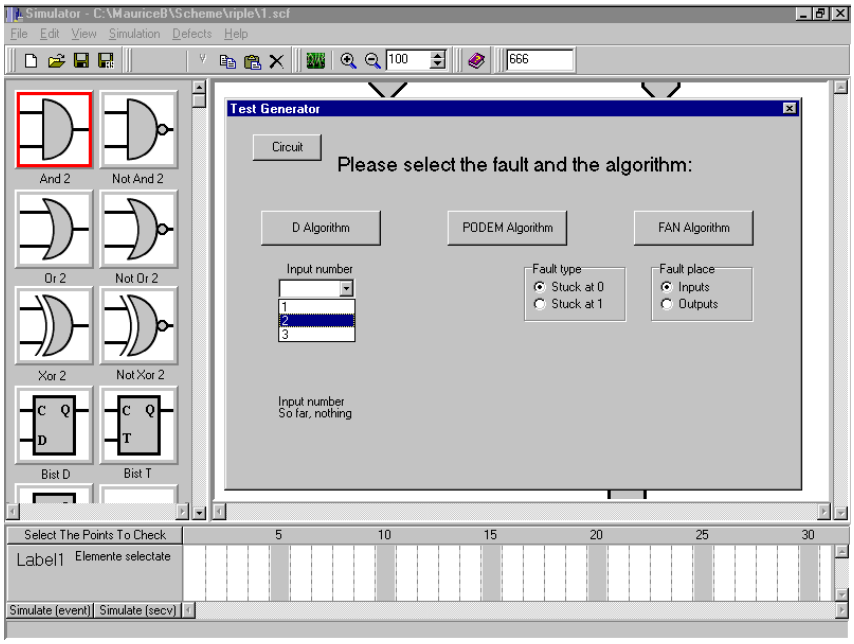


Figure 7:

The design and implementation of TDL software environment includes the software support capable of an optimal generation of stimulus test vectors, optimal simulation and testability assessment, and if it's necessary any modification at structure level in order to increase the digital circuit testability, conforming with [5], [4].

From the following figure (Figure8), we can notice that the user interface is divided in 3 zones:

1)The zone from where we can select the components that will be introduced in the scheme 2)This is practically the zone where the scheme will be edited. Here we will introduce the components, make the connections using wires, etc This is the zone used for displaying the results of the logical simulation of the current scheme. Figure 9 represents the interface for test generation of the tool we developed. The user can select the type of the fault considered ("0" or "1"), the faulty line and the algorithm used for test generation.

6 Comparative results

The modules for test pattern generation have the capabilities to measure the number of backtrackings and the amount of memory that is involved in the generation of the test pattern for a given fault. In order to compare the effectiveness of the three algorithms presented above, we have considered for testing the Ripple Carry Adder (Table 1) and the . Riiple Carry Lookahead Adder (Table 2).

We have generated the test patterns for all the single faults affecting the adders' lines and we have calculated the average number of the backtracking that occurred. We have also calculated the average amount of memory implied for running these modules.

The simulations were made for a duration of 100.000 ns, on a Pentium 4 2 GHz, 1Gb Ram, Windows XP operating system

Table 1

Range of adder	Algoritm D		PODEM		FAN	
Range of adder	Average of Backtrackings	Memory involved	Average of Backtrackings	Memory involved	Average of Backtrackings	Memory involved
1	0,375	12kb	0	19kb	0	32kb
2	3,1s	21kb	2.78	45kb	2.56	79kb
4	9s	36kb	8.3	106kb	8.01	186kb
8	1,51	75kb	1,47	212kb	1,22	378kb
16	2,35	161kb	2,15	433kb	1,86	808kb
32	3,14	328kb	2,56	892kb	2,13	1702kb
64	4.05	665kb	2,87	1824kb	2,38	3492kb

Table 2

Range of adder	Algoritm D		PODEM		FAN	
Range of adder	Average of Backtrackings	Memory involved	Average of Backtrackings	Memory involved	Average of Backtrackings	Memory involved
4(m=2)	0,365	15kb	0.56	23kb	0.41	43kb
8(m=4)	1,56	83kb	1,53	223kb	1,26	395kb
16(m=8)	2,65	179kb	2,4	454kb	1,91	815kb
32(m=16)	3,78	352kb	2,87	904kb	2,15	1998kb
64(m=32)	4.78	710kb	2,99	1925kb	2,54	3686kb

7 Conclusions

These results show that PODEM algorithm reduces the number of backtrackings compare to D algorithm and FAN reduces them even more. The increasing of the logical gates number (5 gates for 1 range adder and 40 gates for 8 range adder) involves a more general view of the algoritm's performances.

References

[1] Abramovici, M., Breuer, M.A., Friedman, A.D., (1996), *Digital Systems And Testable Design*, Computer Science Press

[2] Dotan, Y., Arazi, B., (1990), Concurrent Logic Programming As A Hardware Description Tool, *IEEE Trans. on Comp.*, vol. 39, no. 1, January 1990, pp. 72-88

[3] Fujiwara, Hidao, (1990), Logic Testing And Design For Testability, *Computer Systems Series The MIT Press*, pp.30-75

[4] Popescu D.E., Popescu C., (1994), An Algorithm for Solving the Generalize Probability Problem for Boolean Circuits, *Proceedings of FEI '25 Conference on Electronic Computers and Informatics*, Kosice Herl'any Slovakia, pp.126-132

[5] Popescu C., Popescu, D.E., (2000) Some aspects about applying Boundary Scan Standard, *RSEEE 2000 - Oradea, Computer Science And Reability*, pp.87-93

[6] Popescu, D.E., Popescu C., (1996), Nonrobust path delay fault simulation by parallel processing of patterns, *International Symposium on Systems Theory Robotics, Computers & Process Informatics*, Section Computer Science & Engineering

Daniela E. Popescu, Mirela Pater
University of Oradea
Department of Computer Science
University Street, no.1
E-mail: {depopescu,mirelap}@uoradea.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 460-464

Virtual Heritage Reconstruction Based on an Ontological Description of the Artifacts

Dorin Mircea Popovici, Crenguţa Mădălina Bogdan, Andreea Matei,
Valentina Voinea, Norina Popovici

Abstract: The paper brings into discussion the use of ontologies in the 3D virtual reconstruction of historical sites. As the first attempt, we present the obtained taxonomy, as part of an ontology for a specific historical site and a very specific period. The taxonomy is constructed basis on the top-level ontology DOLCE+D&S. Furthermore, we present the taxonomy utilization in the authoring process of the 3D virtual reconstruction. The technology involved consist in Protégé and RacerPro for ontology, ARéVi as multimodal immersion open-source API, C++/Java as programming languages for system components implementation and XML and VRML as file format for the 3D virtual models and environment description.
Keywords: Virtual heritage, ontology, taxonomy, 3D model

1 Introduction

Application of virtual/augmented reality technologies [14] in the domain of virtual historic visits has become a frequent solution, with the growth of computational power and of the evolution of specific technologies. Now it is possible to visualize immemorial archaeological sites [2] and explore them in detail. However, the senzation of immersion in a virtual environment grows considerably when using a stereoscopic projection CAVE-like system [1].

Populating the environments with virtual humans that simulate people [13] increase the realism of the reconstructed environment by bringing back to ”life” ancient times [9, 6]. Descriptions and population behavior may not always be sufficient to completely describe places and historical periods. It may be necessary to add supplementary information, such as topography, fauna, flora, climate and surroundings, but also multimedia documents (attached to some elements from 3D scenes), real guides in virtual environments, avatars or even virtual guides [10].

Our approach in virtual heritage reconstruction is two steps based: first we construct an historic taxonomy, as part of an domain ontology for describing the context topology and then, basis on this taxonomy, the evolution of the reconstructed environment is described using open-scenarios. In the followings, we are focused on the first issue. Our contribution is organized in two main sections: one dedicated to the ontology itself and the second treats some technical aspects of ontology-based virtual reconstruction. Finally, we present our conclusions and some of our perspectives for the near future.

2 Historic Domain Ontology

An ontology is a formal specification of the concepts intension and the intensional relationships that can exist between concepts. Using logical axioms, it is a declarative model of a domain. In respect to other models, ontologies allow accurate expression of meaning of models.

Ontologies were introduced in the computer science domain over ten years ago and since then they gained an important role in Artificial Intelligence, Computational Linguistics, Database Theory, and Web Semantic.

According to Guarino’s definition, "*an ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world*" [5]. A conceptualization is a set of conceptual (intensional) relations defined on a domain space [5].

Nowadays, there are some top-level ontologies (like DOLCE, SUMO and BFO) which describe very general concepts like space, time, matter, object, event, etc., i.e. independent concepts by a particular domain or problem. Among these, we used the DOLCE ontology [7] and one of its modules D&S [3]. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to particulars. Other top-level ontologies might be used.

In this paper, we construct the taxonomy of the roman artifacts found in Constanta. A taxonomy is a "view" of an ontology, that is, it shows all the concepts of ontology, that are organized after the subsumption relation [7].

Concept name	Informal description
crater	Big dimensions vessel in which the wine and water were mixed
chiton	Classical Greek piece of garment consists of a rectangle piece of cloth, which was draped around the body and caught by an edge and shoulders with fibula
trirema	Big dimensions warship, whose propulsive power was provided by three rows oars. These ships were preeminently used in the II-III centuries B.C.
tiara	A crown-like jeweled headdress

Table 1: Informal description for some important artifacts

2.1 Construction of a Taxonomy of the Roman Artifacts

The methodology of ontology construction basis on the few existent methodologies, like ontology development 101 and others. From these methodologies, we used the method that is presented in [8]. According to this method, in order to construct an ontology we follow the next steps: a) determine the domain and scope of the ontology; b) consider reusing existing ontologies; c) enumerate important terms in the ontology; d) define the classes and the class hierarchy; e) define the properties and relations of classes; f) create instances. For the construction of a taxonomy, the first four steps are necessary.

2.2 Domain and the Scope of the Ontology Identification

The taxonomy models the roman epoch of the Tomis fortress-Constanta, Romania, between the years 46 A.C. and 610 A.C. and the founded objects from that period. We consider both the ships, vessels, constructions types, pieces, and clothing accessories or armament elements of the roman fighters. In addition, we formally described all the concepts that participate in the intension of the concepts above mentioned.

2.3 Identification of the Essential Concepts and Taxonomy Construction

To our knowledge, an ontology of the roman objects has not been constructed until now. Furthermore, we identified the important concepts for the studied domain. In Table 1 we give the informal description of the semantics for some of the most important concepts uses in the virtual reconstruction.

The next step is the definition of the class taxonomy, in which we map each concept, identified in the previous step, in a class and sequent generalization of them. The generalization was done on the basis of the subsumption relation. Furthermore, we made an ontological commitment by using of the top-level DOLCE ontology, along with one of its submodules: D&S.

In the next, we present the taxonomy of the roman objects after the categories that represent ”roots” of their sub-taxonomies. In addition, these categories are directly related by a DOLCE or D&S category. We further mention that in the construction of the taxonomy we used the OWL-DL (Web Ontology Language-Description Logic) language [15] and the Protégé editor [4].

For instance, an important category is that of constructions which were executed by the ancient Romans, like basilica, altar, capitel, etc. and their decorative elements. We present their taxonomy in the figure 1.

2.4 Taxonomy Verification and Validation

In DOLCE, the restrictions are given using a subset of the first-order logic and their verification is a long time task. That is why, we translated our taxonomy in OWL DL language [15] and we checked its consistency with the help of the Protégé tool [4] and the RacerPro reasoner system [11]. Furthermore, the taxonomy has been validated by the National History and Archeology Museum of Constanta, Romania.

3 From taxonomy/ontology to the 3D virtual models

In our approaches, the reconstructed virtual environments (VEs) are defined from the user's perspective. For this, each VE is populated by virtual objects and virtual entities that correspond to objects populating the real world. As presented in [10], by using various criteria, the set of entities within the virtual environment may be

structured in order to obtain complex entities. An entity within the virtual environment, which is able to perceive, decide, and react based on its internal needs, objectives and abilities.

In order to complete the VE description it is necessary to give a more precise description of the semantics of a virtual entity; or, in other words, to formally define the semantics of a virtual entity. To this end, the semantics of each virtual entity is associated with the intension of the corresponding concept in the real world. As we know, in order to model an object in the real world some essential properties of the object should be used [7]. They will make up the object's semantics from the modeler's point of view.

In the virtual environment, for each property of the real object the modeler considers essential will be an attribute of the virtual entity. This way, the obtained taxonomy, based on the real object's properties, gives us two complementary perspectives, one abstract and another physical, on the object's semantics. Later on, this description will be used in VE evolution description/evaluation.

4 Technical aspects

The domain taxonomy is currently completed by the 3D virtual models of the most important artifacts using both 3D laser scanning and 3D virtual reconstruction techniques.

Basis on the produced XML description obtained by the taxonomy writing in OWL DL language, the SceneBuilder prototype allows the interactive authoring of the 3D context. For this, according to the ontology, SceneBuilder present to the user an adaptable to context interface that permits the user to set some physical attributes (as location) of the browsed concept and then, to access one of the 3D models corresponding to the concept, in order to visualize it inside the VE.

This way, depending on the current general context, the user is permitted to add only coherent context. The user actions effect is confirmed by SceneBuilder by adding/updating the concept tree (left side of the figure 2) and by 3D rendering of the artifact instance (right side of the figure 2).

The SceneBuilder output is then passed to an immersive interface, based on the ARéVi open platform developed by CERV [12]. In the figure 3 it is presented a HMD-based user experience using a desktop system with active-stereo visualization. The interface assures to the user a multi-modal immersive experience, where the user is able to explore and to experiment the environment by carrying out complex interactions with the entities and the virtual agents that populates the VE.

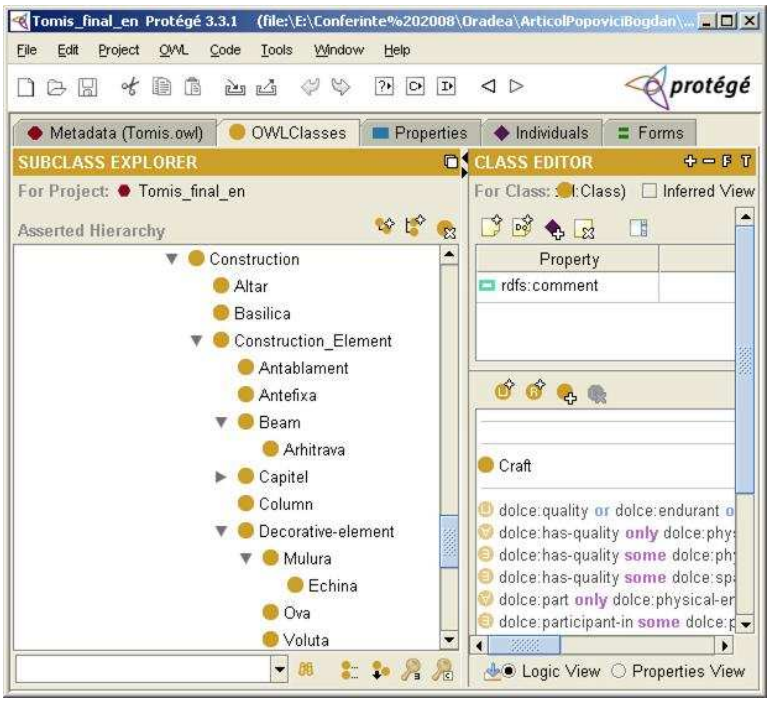


Figure 1: The taxonomy of constructions categories

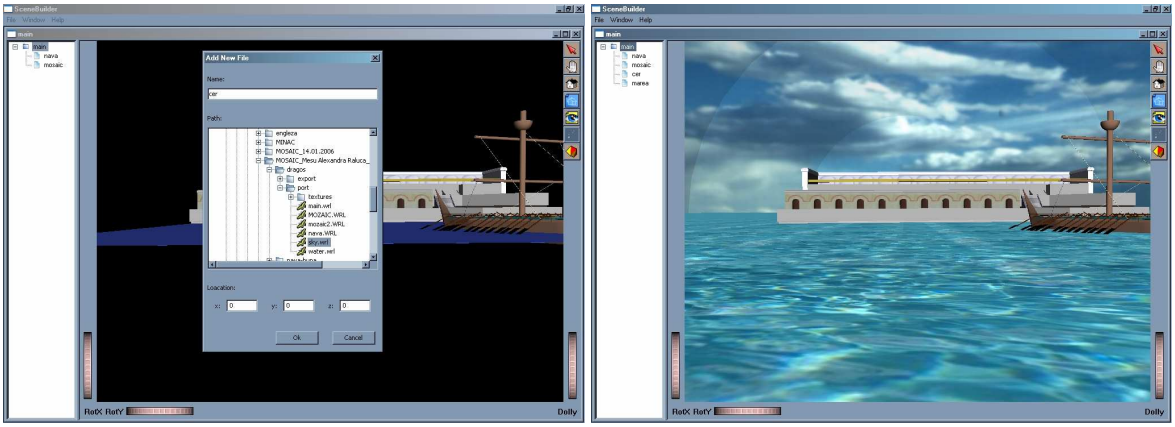


Figure 2: SceneBuilder screenshots

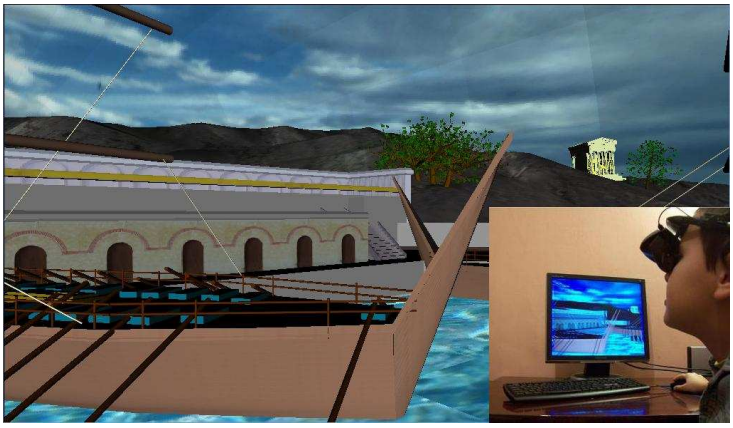


Figure 3: HMD-based immersive experience

5 Conclusions and Future Work

Our contribution demonstrates the use of ontologies in 3D virtual environments authoring assistance and validation tool, taking as concrete example a real historical site. The fact that the author of such a virtual environment has to choose some virtual 3D artifact reconstruction, on the basis of the taxonomy filter, assures (at least) the physical coherence of the proposed 3D virtual reconstructed content. The historical validation of the context is assured by the taxonomy validation.

As we have stated in the introduction, we intend to use our ontology in the behavioral description of the virtual environment also. Therefore, the most promising of our future efforts is oriented in this direction. More, we expect to succeed in offering to the virtual agents that inhabit the virtual environment the ability to reason on the basis of the same ontology. This way, our virtual agents will became more credible.

5.1 Acknowledgements

This work is funded within the TOMIS project, no: 11-041/2007, by the National Centre of Programs Management, PNCDI-2 - Partnerships program. We are grateful to our students Daniel Uzunu for the SceneBuilder prototype development, as well as to Alexandra Meşu and to Dragoş Şerban for their efforts in 3D reconstruction of historical artifacts.

References

[1] C. Cruz-Neira, D. Standin, T. DeFanti, Surround-screen projection-based virtual reality: design and implementation of the CAVE. Proc. of ACM SIGGRAPH'93, pg. 135-142, 1993.

[2] A.G. Gaitatzes, D. Christopoulos, A. Voulgovri, M. Roussou M., Hellenic Culturage Heritage through Immersive Virtual Archeology, Proc. of the 6th International Conference on Virtual Systems and Multimedia, Gifu, Japan, 2000.

[3] A. Gangemi, P. Mika, Understanding the Semantic Web through Descriptions and Situations. In Proceedings of the International Conference ODBASE03, Italy, Springer, 2003.

[4] J. Gennari, M. Musen, R. Fergerson, W. Grosso, M. Crubzy, H. Eriksson, N. Noy, S. Tu, The evolution of Protégé-2000: An environment for knowledge-based systems development. International Journal of Human-Computer Studies, 58(1):89-123, 2003.

[5] N. Guarino, Formal Ontology and Information System. In Proceedings of FOIS'98, Trento, Italy, IOS Press, 1998.

[6] LifePlus - http://lifeplus.miralab.unige.ch/HTML/home.htm

[7] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, WonderWeb Deliverable D18. Ontology Library. IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web, 2003. http://dea.brunel.ac.uk/project/murale/strat.htm

[8] N.F. Noy, D. McGuinness, A Guide to building ontologies: Ontology Development 101. A Guide to Creating Your First Ontology, March, 2001 at http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ ontology101-noy-mcguinness.html

[9] G. Papagiannakis, et al., Mixing Virtual and Real scenes in the site of ancient Pompeii, Journal Of Computer Animation and Virtual Worlds, December, 2004.

[10] D. M. Popovici, et al., Interactive Distributed Guided Tours of Historical Sites, Proceedings of the International Conference on CyberWorlds, pg. 453-457, IEEE Computer Society, ISBN: 0-7695-1922-9, 2003.

[11] RacerPro Reasoner, http://www.racer-systems.com/

[12] P. Reignier, F. Harrouet, S. Morvan, J. Tisseau, T. Duval, ARéVi : a virtual reality multiagent platform, Proceedings VW'98, Paris, july 1-3, Springer Verlag, LNAI1434,pg. 229-240,http://www.enib.fr/ harrouet, 1998.

[13] B. Ulicny, D. Thalmann, Crowd Simulation for Virtual Heritage, Proc. First International Workshop on 3D Virtual Heritage, Geneva, pg. 28-32, 2002.

[14] V. Vlahakis, et al., ARCHEOGUIDE: Challenges and Solutions of a Personalized Augmented Reality Guide for Archaeological sites. IEEE Computer Graphics and Applications (5) 22: 52-60, 2002.

[15] World Wide Web Consortium. OWL Web Ontology Language Reference. W3C Recommendation, 2004.

Dorin Mircea Popovici, Crenguţa Mădălina Bogdan, Andreea Matei
Ovidius University of Constanta, Romania
Computer Science and Numerical Methods Department
124 Mamaia Blvd., 900527
E-mail: {dmpopovici, cbogdan}@univ-ovidius.ro, m_andreea2002@yahoo.com

Valentina Voinea
National History and Archaeology Museum of Constanta, Romania
12 Ovidiu Market, 900745
E-mail: vialia_rahela@yahoo.fr

Norina Popovici
Ovidius University of Constanta, Romania
Faculty of Economical Sciences
1 University Street
E-mail: norinapopovici@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 465-469

Concepts of Graph Theory Relevant to Ad-hoc Networks

M. A. Rajan, M. Girish Chandra, Lokanatha C. Reddy, Prakash Hiremath

Abstract: The issues in Mobile ad-hoc networks (MANETs) always bring the attention of research community. The fundamental issues of connectivity, scalability, routing and topology control in MANETS is worth to study. Graph theory plays an important role in the study of these fundamental issues. This paper highlights the concepts of graph theory that are employed to address these fundamental issues.
Keywords: graphs,connectivity,graph spanners,proximity,MANET

1 Introduction

Ad-hoc networks are decentralized, self-organizing networks capable of forming a communication network without relying on any fixed infrastructure. Each node in an ad-hoc network is equipped with a radio transmitter and receiver, which allow it to communicate with other nodes over wireless channels. All nodes can function, if needed, as relay stations for data packets to be routed to their final destination. A special kind of ad-hoc network is the sensor network where the nodes forming the network do not or rarely moves. Further, the nodes of a sensor network are similar. Salient Features of mobile ad-hoc networks [1] are, 1) Use of ad-hoc networks can increase mobility and flexibility, as ad-hoc networks can be brought up and brought down in a very short time. 2) Ad-hoc networks can be more economical in some cases, as they eliminate fixed infrastructure costs. 3) Ad-hoc networks can be more robust than conventional wireless networks because of their non-hierarchical distributed control and management mechanisms. 4) Because of multi-hop support in ad-hoc networks, communication beyond the Line of Sight (LOS) is possible at high frequencies.

In the study of MANETs two areas are of great importance[3-6]: 1) Understanding the fundamental issues like connectivity, scalability and routing and 2) Network modeling and simulation. Since a network can be modeled mathematically as a graph (there exists a bijection between network topology and graph), the Graph Theory concepts play an important role in analyzing these fundamental issues. Also, the problems associated with ad-hoc networks can be explained mathematically. Going further, graphs can be algebraically represented as matrices and hence the study of the network can be automated through algorithms.

A lot of research about ad-hoc networks is carried out using the mathematical models and their simulation rather than experimenting on real mobile ad-hoc networks. Several issues like node density, mobility of the nodes, link formation between nodes and packet routing between the nodes needs to be simulated. To simulate MANETs concepts of graph theory (particularly random graph theory) are utilized.

The very basic purpose of any network is to facilitate exchange of information between any two nodes. This can happen only when the network is connected. Hence the connectivity is one of the fundamental and most important issues of the MANETs. The important factor, which affects the connectivity, is the transmission range of the nodes and the mobility of the nodes. The majority of research work related to connectivity has been done by considering the static network, wherein nodes will be stationary. Some of the concepts of graph theory that are extensively used to study the connectivity issues are graph spanners, proximity graph sparsifications and spectral graph theory. Another important issue is the *scalability*. *Scalability* is the study of network stability, whenever the number of nodes changes; the topology of the network changes.. This is one of the important issues in ad-hoc networks, because of the mobility of the nodes in the network. Addition of nodes to the network may cause the network be disconnected to start with. This necessitates topology control. Some of the fundamental questions that arise during topology change are how the performance of the network and routing will be affected? A lot of work has been done related to topology control utilizing the graph theory concepts like graph clustering, graph partitioning, and graph evolution[5,12,16].

One of the issues in transporting the data packets among the nodes of a MANET or even to outside is *routing*. The factors which can affect the routing are connectivity, mobility of the nodes and the traffic of the network[3,4,5,6]. Routing protocols in mobile ad-hoc networks are more complex than in static networks. One of the easiest routing techniques is flooding, where packets are simply delivered from source to all its neighbor nodes. The neighbor nodes pass the packets to their neighbor nodes and the process continues until the packet reaches the destination node. But, this affects the throughput of the network as flooding introduces congestion. Hence more optimal routing techniques are needed. A lot of research has been done utilizing the concepts of graph the-

ory in devising routing algorithms; some of the concepts that are explored are sparse graphs, graph planarity and proximity graphs[5,13,15,16].

Based on the discussion so far, it is apparent that the graph theory concepts can play a great deal in the study and design of ad-hoc networks. This paper is an attempt to succinctly capture many such concepts scattered in the literature, outlining their applications in the study of ad-hoc networks. In order to present these, the paper is organized as follows: in Part 2, a very brief introduction of few important terminologies and notations of graph theory is given Part 3 presents the graph theory concepts related to connectivity, routing and topology control issues. Conclusions are given in Part 4.

2 Some Basic Definitions from Graph Theory

A graph consists of number of vertices and edges, where an edge is an association between two vertices. As mentioned earlier, there is a bijection between a graph and a network. With respect to the network, a vertex is a node and an edge is a link between two nodes. Mathematically, a graph G is a triplet consists of Vertex Set *V(G)* , Edge Set *E(G)* and a relation that associates two vertices with each edge. An edge between two nodes *i* and *j* is represented as (*i,j*) and by using usual notation, *E(G)* can be written as $E(G) \subseteq \{(i,j) \mid \forall\, i,j\ \varepsilon\, V \text{ and } (i,j)=(j,i)\}$. Two vertices are said to be adjacent to each other, if there exist an edge between them. Two edges are said to be adjacent to each other, if the one of the end vertex of the edges are same. If each edge of a graph is associated with some specific value (weight), graph is said to be weighted graph. The number of edges associated with the vertex is called degree of any vertex v is denoted by $d(v)$. The *minimum degree* of a graph is the least degree of a vertex of a graph denoted by $\delta(G)$ and the maximum degree of a graph is the *maximum degree* of any vertex of a graph denoted by $\Delta(G)$. A graph G is regular if and only if $\triangle(G) = \delta(G)$. A graph G is said to be connected, if for every pair of vertices *u, v* of G, there exist a path, otherwise Graph is disconnected. A disconnected graph has number of components; each component being a connected graph. A planar graph is a graph which can be embedded in the plane, i.e., it can be drawn on the plane in such a way that its edges may intersect only at their endpoints.A dual graph of a given planar graph G is a graph which has a vertex for each plane region of G, and an edge for each edge in G joining two neighboring regions, for a certain embedding of G.

3 Graph theory concepts related to Connectivity, Routing and Topology Control

This section describes the concepts of graph spanners, proximity of graphs(UDG,NNG,RNG,DT and Voronoi diagram) that can be applied to answer the fundamental issues connectivity, routing and topology control issues one by one.

Graph Spanners[10,15]:. A spanner of a graph is a sub-graph that preserves approximate distances between all pairs of vertices. Formally, given $t \geq 1$, a t-spanner of a graph G is a sub-graph S of G such that for each pair of vertices the distance in S is at most t times the distance in G, t is referred to as the multiplicative stretch factor of the spanner. That is $d_S(u,v) \leq td_G(u,v), \forall\, u,v\, \varepsilon\, V$ and S is called multiplicative t-spanners of G[10]. If $r \geq 0$ and $d_S(u,v) \leq td_G(u,v) + r, \forall\, u,v\, \varepsilon\, V$, S is called additive r-spanner of G and r is called additive stretch factor of G. Several graph spanners may exist for a given graph. So between any two nodes these spanners can be different paths between the nodes.The routing algorithms can use these spanners concept to device efficient algorithms as the spanners provide several alternate paths and also throughput of the network can be increased in case of congestion. By using graph spanners one can determine several graph spanners,which are useful in designing certain class of routing algorithms, study of network clustering, partitioning and network topology control. One of the difficulties in dealing with graph spanners in ad-hoc network is how the algorithm can be made distributed with less complexity? The concept of proximity is also related to graph spanners. A lot of work is done in devising algorithms to construct the graph spanners locally. Graph spanners also finds application in many areas including computational geometry, computational biology, and robotics and distributed computing.

Proximity: With respect to mobile ad-hoc network, the study of proximity graphs plays an important role in topology control, connectivity of the network. Proximity represents the neighbor relationship between nodes. Two nodes are joined by a link if they are deemed close by some proximity measure. This certainly affects the network connectivity. It is the measure that determines the type of a graph that results. Many different measures of proximity have been defined, giving rise to many different types of proximity graphs. One such measure is spanning

ratio. It is defined as $Maximum(\{\frac{\eta(u,v)}{\Omega(u,v}\})$, where $\eta(u,v)$ is the length of the shortest path between two nodes u and v, where the edge length is measured by Euclidean distance and $\Omega(u,v)$ is the direct Euclidean distance. There are different types of proximity graphs available. The important ones are Unit Distance Graph (UDG), Nearest Neighbor Graphs (NNG), Minimum Spanning Trees (MST), Relative Neighborhood Graphs (RNG), Delaunay Triangulation (DT), and Gabriel Graphs[10,15].

Unit Disk Graph (***UDG***): is a graph UDG (V, E) with set of nodes V and a link (i,j) between nodes i and j belongs to link set E if and only if their Euclidean distance is less than 1. UDG can be directed in which link is also directed. Most of the ad-hoc network modeling uses UDG. A more generalized UDG is one, in which the link between two nodes is possible if their Euclidean distance is less than $\boldsymbol{r}$, where $\boldsymbol{r}$ is the radius of the circular transmission range of the antenna of the mobile node. This model is well suited for ad-hoc network, as it is almost realistic to ad-hoc network, where the geodesic of the transmission range of the radio signals coverage is almost circular. One of the issue related to this models is to find out the minimum threshold range of the transmitters of the nodes so that network is connected.

Nearest Neighbor Graph(NNG)[10]: of a graph *G(V,E)*is a directed graph, denoted by *NNG(V',E')* is a spanning sub-graph of a graph G with V=V' and there exist a directed edge e between any two nodes i and j if and only if node j is the nearest neighbor of node i(the Euclidean distance is the measure for selecting nearest neighbor). This graph gives the all-pair shortest paths of a given network.For the given network, if one can able to find its NNG, packet routing will become simple and straight forward and also the throughput of the network will be increased, because of reduced delay in the packet delivery. But the failure of a link , which belongs to NNG , makes it disconnected. Hence a more generalized NNG is required to make it fault tolerant. NNG can be generalized to include more than one neighbor. The K-NNG is a generalized NNG, which represents the number of edges from any vertex to K nearest neighbors. There can be more than one NNG of a given graph. NNG can also be a undirected graph; is a spanning sub graph of *Minimum Spanning Tree*(MST).MST of a weighted graph G is a spanning tree of G which is acyclic, connected with the property that the sum of the weights of the edges of the resulting tree is minimum. Since the edges of a graph could contain the same weights, any particular graph G could have multiple MSTs. A variation of NNG is ***Relative Neighborhood Graph*** (RNG); of a graph G (V, E)is a graph RNG (V, E')in which there exist an edge between nodes i and j and if for all nodes k in V, $d(i,j) \leq Maximum(d(u,k),d(k,j))$.

Gabriel Graph (GG)[15]: is a graph which contains a link between nodes i and j if a disk with radius $\frac{ij}{2}$, contains no other node. Formally GG (V, E)is node set V and link set E , such that there exist a link between nodes i and $j$if and only if $d(i,j) \leq \sqrt{d(i,k)^2 + d(k,j)^2}$, $\forall\, k\, \varepsilon\, V,\ k \neq i\, and\, k \neq j$. A Gabriel Graph can be constructed in time $O(n\log n)$ by first finding the Delaunay Triangulation and Voronoi Diagram for the set of points. Then for each edge in the triangulation, if the edge intersects its Voronoi edge, it is added as an edge to the GG.

Voronoi Diagram: Given a set $P = \{P_1, P_2, ..., P_n\}$ of n points in ε, where $\varepsilon = E^m, E\, is\, an\, edge\, set$ in an affine space, it is often useful to find a partition of the space into regions each containing a single point of P.The Dirchlet-Voronoi diagram *V(P)* of P is the family of subsets of ε consisting of the sets $V_i = \cap_{j \neq i} H(p_i, p_j)$ and all of their intersections.Dirchlet-Voronoi diagrams are also called Voronoi diagrams, Voronoi tessellations or Theissen polygons. ***Delaunay Triangulations (DT):*** A very interesting undirected graph can be obtained from the Voronoi diagram is : The vertices of this graph are the points p_i(each corresponding to a unique region of *V(P)*), and there exists an edge between p_i and p_j if and only if, the regions V_i and V_j share an edge. The resulting graph is called a Delaunay triangulation(DT) of the convex hull of P. A triangulation T is called a DT. if and only if a circle which contains any triangle of T, whose vertices fall on the circle's edge, does not contain any other points in it's interior. Figure 1 describes the Voronoi diagram and its DT. DT is used very well in one of the routing protocol called Location Aware Routing(LAR).LAR is a simple Greedy algorithm in which packet is forwarded along the route which uses only DT(Here the neighbor node is chosen for forwarding the packet, which is geometrically nearest to the destination node in a DT). Note that $NNG(V) \subseteq MST(V) \subseteq RNG(V) \subseteq GG(V)\, DT(V)$. There are few interesting properties involving Delaunay Triangulations.

- The Delaunay graph of a planar node set is a planar graph and any angle-optimal triangulation of a node set P is a DT
- Any DT of P maximizes the minimum angle over all triangulations of P and DT is also a dual graph of the Voronoi diagram. It has an edge between any two Voronoi cells which share a Voronoi edge.

DT can be created by creating a Voronoi diagram of P, then creating the dual of this diagram and also using randomized algorithm with complexity $O(n\log n)$. The steps involved in creating it is as follows

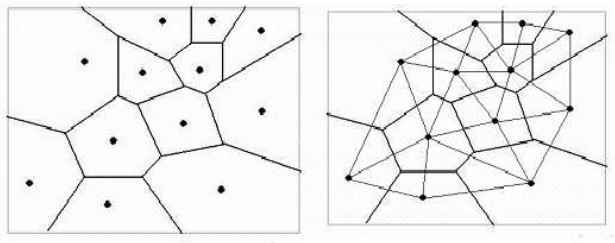


Figure 1: Voronoi diagram and its DT

1. Begin with an initial large triangle, which covers all nodes in set *P*
2. Incrementally insert all nodes one by one while maintaining DT properties explained above.
   - Insert a point into an already legalized DT
   - Triangulate by adding 2 or 3 edges from this node
   - Legalize all possible illegal edges recursively

   Repeat steps until all nodes in set P have been triangulated
3. Remove the initial large triangle

There are lots of applications of Voronoi diagrams and Delaunay triangulations like ; Finding the 1. NNG of a given graph and 2. MST

Restricted Delaunay Graph (*RDG (G)*): It is well known that *DT(G)* is a spanning sub-graph of complete graph, but cannot be constructed locally and may contain long edges. To over come these disadvantages RDG concept is evolved. A RDG is a graph that contains all the short edges of *G* , which is planar and also it is defined as Euclidean spanner of UDG. RDG finds applications related to routing like face routing methods[17] and also in memory less routing algorithm that combines greedy forwarding and local minimum recovery based on face routing. Face routing algorithms applicable only on planar graphs. It is enough, by devising an algorithm to convert the given graph (non-planar) into planar graph by eliminating (or dropping) some links of the graph. In short graph theory plays a vital role in determining, whether a network is planar or non-planar and also to convert non-planar graphs into planar graphs.

4 Conclusions

In this paper we have presented the importance of graph theory to address the fundamental issues in MANET. The paper focuses on the key issues(connectivity,routing,topology control,scaling) and brings an insight of graph theory concepts like graph spanners and proximity graphs . We have extensively studied the graph sparfications,spectral graph theory,random graph models, graph coloring concepts, graph partitioning to address the issues in MANETS[19].

References

[1] ChipElliot and Bob Heile Self-Organizing, Self-Healing Wireless Networks Technical report, BBN Technologies, Cambridge, MA, 2001.

[2] C. F. Huang, Y. C, Tseng, S. L. Wu, and J. P. Sheu, "Increasing the throughput of multihop packet radio networks with power adjustment," *Procedings of International conference on Computer Communications and Netwoks*, pp. 220-225, 2001.

[3] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, in Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking(Mobicom98), ACM, October 1998.

[4] Christian Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in Proceedings of the 3rd ACM international symposium on Mobile Computer Science Department, UCLA, 1976.

[5] G. Di Battista and R. Tamassia. On-line Maintenance of TriconnectedComponents with SPQR-trees.Algorithmica,15:302-318, 1996.

[6] T. K. Philips, S. S. Panwar, and A. N. Tantawi, "Connectivity properties of a packet radio network model," IEEE Trans. Inform. Theory, vol. 35, pp. 1044-1047, Bollob´as, B., Erdos, P.,

[7] Graphs of extremal weights, Ars Combinatoria 50 (1998),225-233, 1989.

[8] F. Harary, Graph Theory, Narosa Publishing House

[9] Charles E. Perkins. Ad-hoc networking. Addison-Wesley, Boston, 2001.

[10] D. Peleg and A. A. Sch¨affer, Graph Spanners, Journal of Graph Theory,13:99-116, 1989.

[11] E. W Dijkstra, A note on two problems in connexion with graphs, Numerische Mathematik, 1959.

[12] M. D. Penrose. Random Geometric Graphs. Oxford University Press, 2003.

[13] O. H¨aggstr¨om and R. Meester, "Nearest neighbor and hard sphere models in continuum percolation," Random Structures and Algorithms, vol. 9, pp. 295-315, 1996.

[14] Dragos M. Cvetkovic Michael Doob Horst Sachsp, "Spectra of Graphs, Theory and Application", Academic Press

[15] X.-Y. Li and I. Stojmenovic, Partial Delaunay triangulation and degree limited localized Bluetooth scatternet formation, in: Proc. AD-HOC NetwOrks and Wireless (ADHOC-NOW), Fields Institute,Toronto, 2002.

[16] R. Albert and A. L. Barab´asi, Statistical mechanics of complex networks, Rev. Modern Physics, 74, 47-97, 2002.

[17] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. Wireless Networks, 7(6):609-616, 2001.

[18] M.A.Rajan, M.Girish Chandra, Lokanatha C. Reddy and Prakash Hiremath, A Study Of Connectivity Index of Graph Relevant to Adhoc Networks, IJCSNS VOL.7 No.11, November, 198-204, 2007.

[19] M.A.Rajan, M.Girish Chandra, "A Study of Graph Thoery Concepts Relevant to MANETS", Technical Report, TCSL, October 2006.

M.A.Rajan[1], M.Girish Chandra[2],Lokanatha C. Reddy[3] and Prakash Hiremath[4]
Research Scholar Dravidian University(DU)and Tata Consultancy Services Limited(TCSL)[1], Consultant TCSL[2], Professor DU[3], Professor, Gulbarga University[4]
Department Of Computer Science[1],Embedded Systems Lab[2]
Kuppam, Andhra Pradesh and TCSL, Bangalore, INDIA[1], Bangalore[2], Kuppam, Andhra Pradesh[3], Gulbarga, Karnataka[4]
E-mail: rajan.ma@tcs.com[1], m.gchandra@tcs.com[2], lokanathar@yahoo.com[3], hiremathps@hotmail.com[4]

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 470-474

Tree-Like Bayesian Network Classifiers for Surgery Survival Chance Prediction

Beáta Reiz, Lehel Csató

Abstract: Bayesian Networks encode causal relations between variables using probability and graph theory. We exploit the causal relations to detect dependency structures in database consisting of a small number of observations having high dimensions.
Bypass surgery survival chance must be inferred from a database consisting of 66 medical examinations for 313 patients. Tree-like Bayesian were inferred based on mutual information and than analysed for classification of data, respect to survival. Bayesian Network approach allows us to interpret the predictions of the system thus helping the doctor in after surgery treatment prescription. In this paper we present the used methods and results on artificial data.
Keywords: Bayesian Network, medical data classification, causal discovery

1 Introduction

In this paper we analyze tree-like Bayesian Network implementations of bypass surgery survival prediction. We aim to establish causal relationships between variables representing medical examinations. The aim is to predict the survival of a particular patient [4] and to obtain the "most relevant" variables affecting the output of the classifier.

We aim to reduce the exponential complexity of the problem by constructing a tree representing the *immediate causal* relationships between class variable and observations [11]. A tree-like Bayesian network structure was inferred from the data, where the root of the tree is the class variable and remaining nodes are attributes.

In a formal way, we have to compute – from the inferred network – the following probabilities:

1. the surgery survival chance:
$$P(Y|\boldsymbol{X}=\boldsymbol{x}) \tag{1}$$
2. the impact of examinations on survival chance:
$$P(X_i = x_i|\neg Y) \qquad i=\overline{1,n} \tag{2}$$

where $Y=\{T,F\}$ is the class variable and $\boldsymbol{X}=[X_1,X_2,\ldots,X_n]^T$ is the vector of attributes, with corresponding values $\boldsymbol{x}=[x_1,x_2,\ldots,x_n]^T$. In what follows we use the vector notation where boldface denotes the vector and we refer to the elements by a simple index.

2 Bayesian Networks

Bayesian networks (BNs) [9, 10] are triplets $(\mathrm{V},\mathrm{E},\mathcal{P})$, where (V,E) is a directed acyclic graph (DAG) with nodes V, edges E and there is a set of probability distributions $\mathcal{P}$, called parameters, whose elements are assigned to the nodes of the graph. The nodes represent domain variables and edges mark direct causal relations between these variables.

The network encodes a joint probability distribution function representative to the domain:

$$P(X)=\prod_{i=1}^{n} P(X_i|\mathrm{par}(X_i))$$

where n is the number of domain variables, X_i is a node from the BN and $\mathrm{par}(X_i)$ is the set of X_i's parents. The aciclicity of the graph ensures the product to be finite.

We will use the following notations: X and Y are random variables, defined on probability spaces Ω_X respective Ω_Y, with the corresponding distribution functions $p(x)$ respective $p(y)$, and their joint and conditional probability function $p(x,y)$ respective $p(x|y)$.

Information theory offers us numerical characterization of uncertainty in domain variables [3]. We are interested in the amount of information two variables X and Y contain about each other, respect to the set of variables Z. This is called conditional mutual information, and can be defined as:

$$I(X,Y|Z) = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} p(x,y|z)\ \log\frac{p(x,y|z)}{p(x|z)p(y|z)}$$

In next section we present the two-phase tree-like Bayesian Network structure learning algorithm developed for the bypass problem, where direct causal relations encoded by the BN are interpreted as the maximum of conditional mutual information [1, 2, 8] between nodes.

3 Network topology learning

We construct the network topology with a two-phase algorithm. In the first phase we're searching for direct dependencies between attributes and class variable, this way constructing a Naive Bayesian network. The second phase consists of applying Chow-Liu's algorithm to learn the inner structure of network and reveal attribute-attribute correlations.

Naive Bayes classifiers [6, 7] are widely used in classification problems. They are called Naive because of the independence assumption of the attributes. Although this is strong assumption when facing real datasets, the Naive Bayes classification is a powerful tool for its simplicity and often gives convenient results. Figure 1. illustrates a Naive Bayesian network.

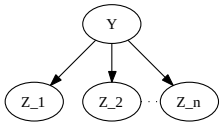


Figure 1: Naive Bayes Net

During the Naive Bayesian network learning process direct dependencies between class variable and attributes has to be find. Dependency relations are interpreted as class variable specifiers. Direct dependences were learned using conditional mutual information maximisation [5], such a way getting a non-redundant Naive Bayesian network. We introduce a threshold parameter – denoted with α_1 – which is the minimum "information" required when putting a new attribute in the network, this way controlling the direct causal relations of class variable and attributes.

Consider the Naive Bayes network from Figure 1. The network is formed of class variable Y respective variables $\boldsymbol{Z}$ directly linked to the class variable. There are also attributes $\boldsymbol{X}$ excluded from the network. We use mutual information maximisation to discover the causal relations between attributes from the network and excluded attributes. We are searching for the excluded attributes that carry almost the same information about class variable as the one already placed in the network. The algorithm is presented below:

Algorithm 1: Tree-like Bayesian Network structure learning

1: place the class variable Y in the network
2: $\boldsymbol{Z} = \emptyset$
3: {*Naive Bayesian structure learning*}
4: **WHILE** $I(X,Y|\boldsymbol{Z}) \geq \alpha_1$
5: $\quad \hat{X} = \underset{\boldsymbol{X}}{\operatorname{argmax}}\, I(X,Y|\boldsymbol{Z})$
6: $\quad$ place $\hat{X}$ in the network
7: $\quad \boldsymbol{X} = \boldsymbol{X} - \{\hat{X}\}, \boldsymbol{Z} = \boldsymbol{Z} \cup \{\hat{X}\}$
8: {*Inner structure learning*}
9: **WHILE** $\boldsymbol{X} \neq \emptyset$
10: $\quad \left[\hat{X},\hat{Z}\right] = \underset{X_i,Z_j}{\operatorname{argmax}}\, I(X_i,Z_j)$
11: $\quad$ place edge between $\hat{X}$ and $\hat{Z}$
12: $\quad \boldsymbol{X} = \boldsymbol{X} - \{\hat{X}\}, \boldsymbol{Z} = \boldsymbol{Z} \cup \{\hat{X}\}$

The threshold parameter α_1 assures the selection of relevant attributes respect to the class variable. The result is a tree-like Bayesian network as in Figure 2, where the root of the tree is the class variable, and the other nodes

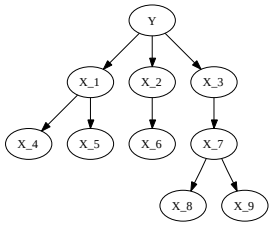


Figure 2: Possible structure of learned BN

are attribute variables. The orientation of edges is from parent to the child, this way minimizing the modification of network parameters during a learning step.

The above algorithm is deterministic, and since we are unsure about domain dependency relations and the accuracy of parameter estimation importance sampling [12] was introduced in the objective maximization function. The distribution used for sampling is a transformation of mutual information to a distribution function.

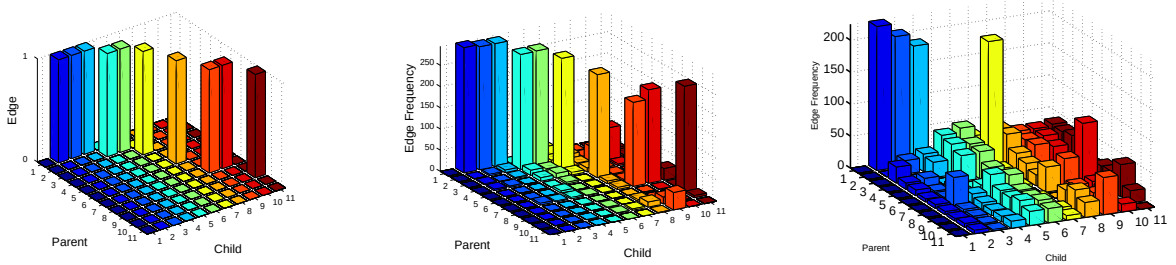
We used two transformations during the tests. The first – denoted f_1 – is the normalization of the mutual information:

$$\mathrm{f}_1(X) = \frac{I(X,Y)}{\sum_{X' \in \boldsymbol{X}} I(X',Y)} \tag{3}$$

The second function – denoted f_2 – uses the exponentiation transform of the mutual information. It has a β parameter which can be understood as a temperature parameter and it controls the constructed distribution function. The higher this parameter is the higher is the probability of selecting the maximum of mutual information.

$$f_2(X) = \frac{\exp(1+\beta \cdot I(X,Y))}{\sum_{X' \in \boldsymbol{X}} \exp(1+\beta \cdot I(X',Y))} \tag{4}$$

Figure 3. shows the histogram of learned edges using the presented approaches and also the generator network structure of data. In each graph on the horizontal plane is the adjacency matrix of the network topology, and the vertical columns represent the histogram edges. We consider the first attribute from the network as the class variable, so it's the root of the constructed tree. Figure 3(a) represents the generator network topology for the data. The other graphs – Figure 3(b), 3(c) – represent the frequency of edges in the learned network topologies during 300 tests.



(a) Generator network topology

(b) Learned topologies with function f_1

(c) Learned topologies with function f_2, $\beta = 20$

Figure 3: Generator network and histogram of BN edges

Although there is a randomness introduced with importance sampling in our first approach, the generated structure is relatively stable when learning with function f_1. The direct causal relations between the class variable and attributes are almost the same during the 300 tests simulations, differences can be observed only on the third level of the tree. There in approximately 100 cases – out of 300 – a single edge is placed differently compared to the generator network.

A much higher degree of randomness can be observed when learning with function f_2 for $\beta = 20$ (Figure 3(c)). One can see that the randomness can be observed mostly when learning attribute-attribute correlations. The

Naive Bayesian networks structure seems relatively stable. Some edges from the inner structure of the BN are also learned with high frequency and they are present also in the generator network. All non-stable attribute-attribute correlations has in common, that the corresponding edge from generator network appears at least as the third highest frequency from the histogram. This demonstrates the high correlation between topology and parameters of a BN.

4 Inference

To compute the probabilities specified in section 1. we inferred a tree-like Bayesian network. The function $\text{par}(X_i)$ denotes the parent of the node X_i. Then the joint probability distribution encoded by the network is:

$$P(X,Y) = P(Y) \prod_{i=1}^{dim(X)} P(X_i|\text{par}(X_i))$$

The survival chance can be given using Bayes theorem, as the ratio of the joint distribution encoded by the network and the marginal distribution of the attributes:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{\sum_{y \in \boldsymbol{Y}} P(X|Y)P(Y)}$$

The second task of our problem is to compute the impact of each attribute on the class variable's value. The impact of the given attributes can be computed as:

$$P(j := X_i|\neg Y) = \prod_{j}^{j \neq Y} P(j|j := \text{par}(j)) \tag{5}$$

where $:=$ denotes value assignment. So the eq. (5). defines a recursive lifting in the constructed tree, until we reach the class variable. The higher the probability defined by eq. (5). is, the more impact has the respective attribute on class variable. We use the negated value of class variable because of the bypass problem.

5 Conclusions

In this paper we presented a stochastic tree-like Bayesian Network classifier algorithm developed for medical data classification. The algorithm uses mutual respective conditional mutual information maximisation to find direct causal relations.

KL-distance between generator and learned BN is computed each time an attribute is placed in the network, to visualize the convergence of the learning process. The results are shown on Figure 4.

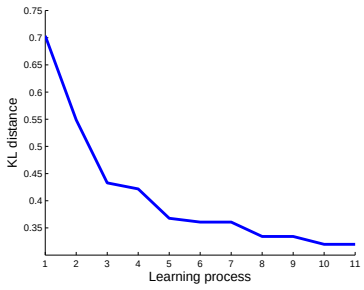


Figure 4: Model fitting during learning

Classifier performance is presented in Table 1. One can see that learning with function f_2 performs better than learning with function f_1. Also for higher β values we get better results.

Before applying the algorithm on real data we have to binarise the observations. We propose binarisation of continuous data with a cutting function based on mutual information for each attribute. We define cutting points,

Tree like BN with function f_1	85.85%
Tree like BN with function f_2, $\beta = 10$	85.88%
Tree like BN with function f_2, $\beta = 20$	85.95%
Tree like BN with function f_2, $\beta = 50$	86.08%

Table 1: Efficiency of presented algorithms

which identifies the binarised value of the respective attribute. The optimal cutting point gives the maximum of mutual information between class variable and respective binarised attribute from all possible cutting point based binarisations.

Relevant topologies and the α_1 parameter will be selected such that the respective BN give the highest performance on a set of testing data.

Acknowledgements

We acknowledge the support of the Romanian Ministry of Education, grant CEEX/1474 and thank for the problem description and the medical database to Béla Vizvári from Department of Operations Research, Eötvös Loránd University, Budapest.

References

[1] J. Cheng, D. Bell, and W. Liu, "An algorithm for bayesian belief network construction from data," 1997.

[2] Jie Cheng, David A. Bell, and Weiru Liu, "Learning belief networks from data: An information theory based approach," *In CIKM*, pages 325-331, 1997.

[3] Thomas M. Cover and Joy A. Thomas, *Elements of information theory.* Wiley-Interscience, New York, NY, USA, 1991.

[4] Zs. Csizmadia and B. Vizvári, "Methods for the analysis of large real-valued medical databases by logical analysis of data," *Rutcor Research Report RRR 42-2004*, Rutgers Center for Operations Research, Rutgers University, 2004.

[5] F. Fleuret, "Fast binary feature selection with conditional mutual information," *Journal of Machine Learning Research*, Vol. 5:1531-1555, November 2004.

[6] Nir Friedman, Dan Geiger, and Moises Goldszmidt, "Bayesian network classifiers," *Machine Learning*, Vol. 29(2-3):131-163, 1997.

[7] David Heckerman and Christopher Meek, "Models and selection criteria for regression and classification," *Technical Report MSR-TR-97-08*, Microsoft Research, 1997.

[8] Mieczyslaw A. Klopotek "Mining bayesian network structure for large sets of variables," *In ISMIS*, pages 114-122, 2002.

[9] David E. Heckermann, *Probabilistic similarity networks,* MIT Press, 1991.

[10] K. Murphy. Learning bayes net structure from sparse data sets. Technical report, Comp. Sci. Div., UC Berkeley, 2001.

[11] Judea Pearl. *Causality: Modeling, Reasoning, and Inference.* Cambridge University Press, Cambridge, 2000.

[12] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

Beáta Reiz
Babeş Bolyai University
Faculty of Mathematics and Computer Science
1 Kogălniceanu str. RO-400084 Cluj-Napoca
E-mail: reiz.bea@gmail.com

Lehel Csató
Babeş Bolyai University
Faculty of Mathematics and Computer Science
1 Kogălniceanu str. RO-400084 Cluj-Napoca
E-mail: csatol@cs.ubbcluj.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 475-479

Visual Based Lane Following for Non-holonomic Mobile Robot

Amar Rezoug, Mohand Said Djouadi

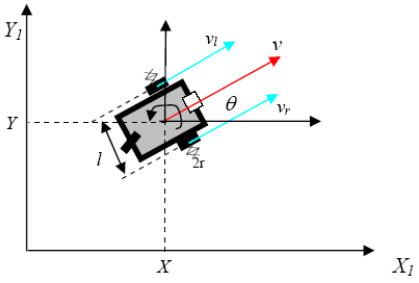
Abstract: This paper aims to contribute in mobile robots endowment with autonomy when accomplishing robotic tasks. The problem treated concerns path following by non-holonomic mobile robot (Pioneer II DX) using visual data coming from CCD camera mounted on the latter. The current advancing path is approximated by a linear function in case of straight line path and by a third function in curve case. The obtained experimental results confirm the effectiveness of the proposed approach.
Keywords: Path following, Visual servoing, trajectory generation, trajectory following.

1 Introduction

In recent years a lot of research effort has been consecrated to mobile robots endowment with autonomy when accomplishing robotic tasks. This might provide many advantages, especially to decreased costs in automated factories. In this paper the problem of automated lane-following for mobile robots is considered; this consist in an autonomous mobile robot which is in charge of following the path traced by a lane marked, out on the floor. It would be interesting to precise that one major application field for the lane-following problem that of automated highways. Path or road-following using visual sensors is a relatively well developed area of research with several examples of vehicles capable of covering thousands of kilometers across roads [1,2]. The paper is organized as follows. Section II presents the kinematical model of the considered mobile robot (PioneerII DX). Section III discusses the image processing techniques and approaches used to extract the needed visual data. Section IV is devoted to the inverse and perspective point transformation. Section V analysis the trajectory estimation in the real world, and section VI describes the steering control approach. Finally, experimental results are presented and discussed in section VIII.

2 Kinematical Model Of Mobile Robot

The robot considered in this work is a two-wheeled one, it is the pioneer II DX manufactured by Activemedia corporation. Its wheel rotation is limited to one axis. Therefore, the navigation is controlled by the speed change on either side of the robot. It has non-holonomic constraints, which should be considered during path planning. The kinematical scheme of our mobile robot is described in Figure 1, where v is the velocity of the robot,v_l is the velocity of the left wheel, v_r is the velocity of the right wheel, r is the radius of each wheel, l is the distance between two wheels, X and Y are the position of the mobile robot, and θ is the orientation of the robot.



(a) Kinematical scheme of the mobile robot

(b) Mobile robot Pioneer II DX used in our experimentations.

Figure 1: Kinematical scheme of the Pioneer II DX

We suppose, as it is usually done, that the contact between the wheel and the ground satisfies both conditions of pure rolling and nonslipping during the motion; moreover, the robot is assumed to be rigid. According to the motion principle of rigid body kinematics, the motion of a mobile robot can be described using equations (1) ,

where ω_l and ω_r are angular velocities of the left and right wheels respectively, and ω is the angular velocity. The left and the right velocities of the robot are:

$$v_r = r \cdot \omega_r, \qquad v_l = r \cdot \omega_l, \qquad \omega = \frac{v_r + v_l}{l}, \qquad v = \frac{v_r + v_l}{2} \tag{1}$$

Combining the set of equation(1) we obtain:

$$\omega = \frac{r}{l}(\omega_r - \omega_l), \qquad v = \frac{r}{2}(\omega_r - \omega_l) \tag{2}$$

The dynamic function of the robot is defined as:

$$\omega_r = \frac{1}{r}v + \frac{l}{2r}\omega, \qquad \omega_l = \frac{1}{r}v + \frac{l}{2r}\omega \tag{3}$$

Finally the mobile robot model is:

$$\dot{X} = v \cdot \cos\theta, \qquad \dot{Y} = v \cdot \sin\theta, \qquad \dot{\theta} = \omega \tag{4}$$

Equation (3) and (4) describe the kinematical model of a two-wheeled mobile robot. While the control variables are the angular velocities of the left and the right wheels. The model also shows that it is about a non-linear system.

3 Image Processing

There exist a lot of image-processing algorithms extracting a map of the environment from the data provided by a camera. Therefore, the implementation of such sophisticated algorithms is usually difficult [4]. Several algorithms and techniques of lane detection are usually used. Among of them, we can name: The Hough transform in [1] fuzzy logic in [1, 5] etc. We can classify the detection of lane in to two approaches: 1/ *Approach based on lane features*: localize in the picture the lines painted on the road, while combining low level techniques as contour detection and the traditional segmentation. This approach requires that the road would be marked correctly by visible lines otherwise the detection would be very difficult and quiet impossible. 2/ *Approach based on a lane model*: consider the problem of lane detection like a problem of estimation, where it is necessary to estimate a set of parameters of a lane model. In our experimental case, we have privileged this approach. The *CCD* Pan-tilt camera mounted on the mobile robot allows acquisition of the lane image (Fig. 2). In order to extract the desired data from the captured images, particularly the lane shape, we used the following image processing algorithm which we can summarize in four steps: Step1: Image acquisition Step2: *YCbCr* image transform (*YCbCr* family of color spaces used in video systems). Y is the luminance component, *Cb* and *Cr* the chrominance components. *YCbCr* is sometimes abbreviated to *YCC* [6],*YCbCr* signals are created from the corresponding gamma adjusted *RGB* source using two defined constants k_b and k_r as follows:

$$\begin{aligned} Y &= k_r \cdot R' + (1 - k_r - k_b) \cdot G' + k_b \cdot B' \\ Cb &= \frac{0.5 \cdot (B' - Y)}{(1 - k_b)} \\ Cr &= \frac{0.5 \cdot (R' - Y)}{(1 - k_r)} \end{aligned} \tag{5}$$

Where $k_b$and k_r are derived from the definition of the *RGB* space. R',G' and B' are assumed to be nonlinear and nominally range from 0 to 1, with 0 representing the minimum intensity and 1 the maximum. Mathematical morphology is a useful tool for image segmentation and processing. Due to the complexity of color-scale morphology, we transform color image into binary image by subsampling techniques and threshold-based segmentation. Then operators in the area of binary morphology including dilation, erosion, closing and opening are used for data preprocessing [7]. Then steps 3 and 4 are: Step3: Image *Cb* plan binarization Step4: Opening and closing operators [7], a closing operation is defined as: $A \bullet I = (A \oplus I) \Theta I$ an opening operation is defined as: $B \bullet I = (B \Theta I) \oplus I$. Where: I is the image matrix, A and B are operators. Finally, we extract the image coordinates of a lane point to be tracked. The latter, could be chosen every where on the lane, in our case we selected whose y coordinate is null.

(a) Original image (b) Image after processing

Figure 2: Original and processed image

4 Inverse and Perspective Point Transformation

A. *Perspective Point Transformation*: Assuming that the road surface is plane, a three-dimensional robot coordinate system is defined as shown in figure (3-b). We suppose that the position of the lane point P is expressed by the coordinate (X',Y',Z') in the robot coordinate system (RCS), the coordinate (x,y) of the corresponding point in the image plane is expressed by the following formulas.

$$x = f\frac{X'}{Y'} \qquad y = f\frac{Z'}{Y'} \tag{6}$$

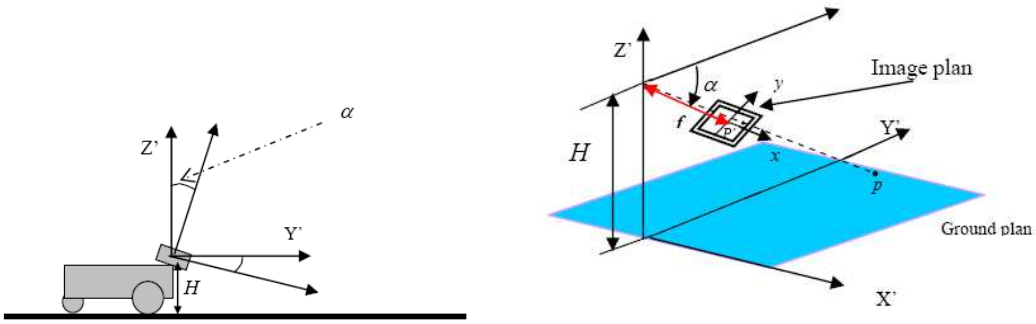
In case tilt inclination of the camera by an α angle rotation figure (3-a). Assume the centre point of camera's lens is the origin of the RCS. The perspective relationship between RCS and the image coordinate system (ICS) is shown at figure (3-b). The perspective point $p'(x,y)$ of a static point $p(X',Y',Z')$ in ICS can be transformed by:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\sin\alpha & \cos\alpha \\ 0 & \cos\alpha & \sin\alpha \end{bmatrix} \tag{7}$$

$$x = X', \qquad z = Y'\cos\alpha + Z'\sin\alpha \qquad y = -Y'\sin\alpha + Z'\sin\alpha \tag{8}$$

$$x = \frac{f\cdot X'}{Y'\cdot\cos\alpha - Z'\sin\alpha}, \qquad y = \frac{(Y'\cdot\cos\alpha + Y'\sin\alpha)}{Y'\cdot\cos\alpha - Z'\sin\alpha} \tag{9}$$

Where $f = 54mm$ camera's focus and α is the tilt angle of the mounted camera. There is a built-in assumption that all these state parameters of the camera are known and the road surface is plain.



(a) Camera inclination angle (b) Relationship between *ICS* and *RCS*

Figure 3: Relationship between *ICS* and *RCS*

B. *Inverse Point Transformation*: A bird-eye view of lane is obtained by the inverse perspective transform of image data to real 3D space coordinates using the perspective relationship between two coordinate systems. If the

elevation of the camera is known, the points on the road surface can be calculated by:

$$X' = Z' \cdot \frac{f \cdot (\cos\alpha + y \cdot \sin\alpha)}{y \cdot \cos\alpha - f \cdot \sin\alpha}, \qquad Y' = Z' \cdot \frac{x}{y \cdot \cos\alpha - f \cdot \sin\alpha}, \qquad Z' = -H \tag{10}$$

Where H indicates camera's elevation above the road surface, (x,y) is the image coordinate in *ICS*, and (X',Y',Z') is the 3-D coordinate in *RCS*.

5 Trajectory Models in the RCS

To approximate the trajectory in the RCS, two cases are taken into account: *first case*: when the portion of the global trajectory to be followed is curve. Here a third function is chosen:

$$Y' = A \cdot X'^3 + B \cdot X'^2 \tag{11}$$

With: $A = \frac{X'\tan\theta_1 - 2Y'}{X'^3}$ and $B = \frac{3Y' - X'\tan\theta_1}{X'^2}$
Where θ_1 is the orientation of point P. *second case*: when the portion of the global trajectory to be followed is straight line. Here a linear function is chosen, because as we can remark in the equations above in case $\theta_1 = 0$ it is very difficult to approximate the trajectory by third equation for this reason we choose the function: $Y' = CX' + D$

6 Steering Control

The Steering Control is determined by means of a virtual reference line. The control law is defined as follows: -The forward velocity is imposed to be constant: $v = v_0$ -The angular velocity: $\omega = k \cdot B \cdot v$ while v is translation velocity. k is an experimental coefficient determined by tests and B is determined by the equation above. the case $\theta = 0$ (straight line) the translation velocity increases and angular velocity equal zero: -The translation velocity is imposed to be constant:

$$v = V \,(v_0 \prec V) \tag{12}$$

-The angular velocity:

$$\omega = 0 \qquad \text{In case} \qquad \theta_1 = \frac{\pi}{2} \Rightarrow \tan\theta_1 \to \infty \tag{13}$$

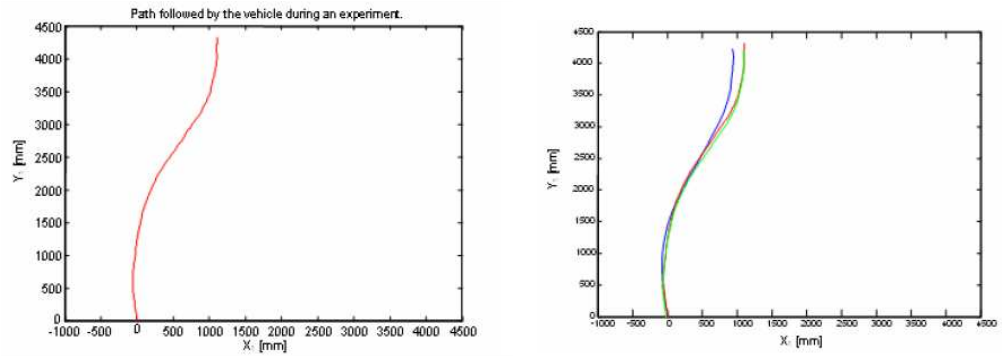
the robot moves arbitrarily until acquiring a new lane point with $\theta_1 \neq \frac{\pi}{2}$

7 Visual Based Lane Following Algorithm

When the robot drives from its actual position to its future one, it executes the following algorithm: **Begin Repeat** 1/ Video acquisition and images processing./2 A target point is chosen among the lane points. 3/ Inverse point transformation, given by equation (10). 4/ Trajectory estimation in the *RCS* using linear function in case of straight line trajectory or third function else. 5/ Steering control.
End.

8 Experimental Results

For experimental evaluation of the proposed control algorithm, we used the mobile robot Pioneer II DX figure(1-b). The latter is equipped by its own controller. The vision system includes a frame grabber PXC200 that allows capturing images thanks to a camera SONY EV-D30 mounted on the robot. These images are radio transmitted from the robot to a Pentium IV 2.4 Ghz PC, in charge of processing the images and computing the corresponding control actions. Figure 2-a shows the image captured by the camera; this image is processed into the image result shown in Figure 2-b. From this image, the inverse projection and trajectory generation are calculated and used to compute the variables used by the controller. Finally, the computed control actions are sent by a radio transmitter to the robot.



(a) An example of an experimental lane followed by the mobile robot

(b) Paths followed by the vehicle during different experiments with different light conditions. The path has been reconstructed by means of the sole odometry.

Figure 4: Experimental results

9 Conclusion

This paper proposes a lane following visual based algorithm for a mobile robot. This algorithm is mainly divided into two parts, the first one concerns image processing which provides the robot by the real world coordinates of the point to have to reach, the second one concerns the generation of a virtual trajectory in the robot coordinate system to have to follow and then computes the angular and linear velocities required to achieve like a task. The experimental results seem to be very interesting.

References

[1] Z. Hu and K. Uchimura, "Action-Based Road Horizontal Shape Recognition" *SBA Controle and Automação* Vol. 10 no. 02 / Maio, Jun., Jul. e Agosto de 1999.

[2] Y. Xu, K. Li, Yingma, Yuanyi, Wanjian, Chenjun, Z Yufan, "General Design of the Lateral Control System Based on Monocular Vision on THASV-I" *2004 IEEE Intelligent Vehicles Symposium University of Parma Parma*, Italy June 14-17,2004.

[3] X.jiang, y motai, x. Zhu, "prediction fuzzy logic controller for trajectory tracking for mobile robot" *2005 IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications Helsinki University of Technology*, Espoo, Finland, June 28-30, 2005.

[4] J-B. Coulaud, G. Campion, G. Bastin, and M. De Wan, "Stability Analysis of a Vision-Based Control Design for an Autonomous Mobile Robot" *IEEE TRANSACTIONS ON ROBOTICS*, VOL. 22, NO. 5, OCTOBER 2006

[5] G. Antonelli, S. Chiaverini, "Experiments of fuzzy lane following for mobile robots " *Proceeding of the American control conference BostonP*, MASSACHUSETTS, AACC 2004.

[6] Rodrigo Montúfar-Chaveznava, Fernando Hernández Gallardo and Saúl Pomares Hernández, "Face Detection by Polling" ,*IEEE 2005 Faro*, Portugal,1-3 September 2005.

[7] Z. Hai-bo, Y. Kui, L. Jin-dong, "A Fast and Robust Vision System for Autonomous Mobile Robots" *Proceedings of the 2003 IEEE International Conference on Robotics,Intelligent Systems and Signal Processing* , Changsha, China - October 2003.

Mohand Said Djouadi, Amar Rezoug
Military Polytechnic School
Robotic Laboratory
BP 17, Bordj el Bahri, Algiers Algeria
E-mail: msdjouadi@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 480-485

Simulating NEPs in a cluster with jNEP

Emilio del Rosal, Rafael Nuñez, Carlos Castañeda, Alfonso Ortega

Abstract: This paper introduces *jNEP*: a general, flexible, and rigorous implementation of NEPs (the basic model) and some interestenting variants; it is specifically designed to easily add the new results (filters, stopping conditions, evolutionary rules, and so on) of the research in the area. jNEP is written in Java; there are two different versions that implement the concurrency of NEPs by means of the Java classes *Process* and *Thread*s respectively. There are also extended versions that run on clusters of computers under JavaParty. *jNEP* reads the description of the currently simulated NEP from a XML configuration file. This paper shows how *jNEP* tackles the SAT problem with polynomial performance by simulating an ANSP.
Keywords: NEPs, natural computing, simulation, clusters of computers

1 Introduction

1.1 NEPs

NEP stands for *Network of Evolutionary Processors*. NEPs are an abstract model of distributed/parallel symbolic processing presented in [1, 2]. NEPs are inspired by biological cells. These are represented by words which describe their DNA sequences. Informally, at any moment of time, the evolutionary system is described by a collection of words, where each word represents one cell. Cells belong to species and their community evolves according to mutations and division which are defined by operations on words. Only those cells are accepted as surviving (correct) ones which are represented by a word in a given set of words, called the genotype space of the species. This feature parallels the natural process of evolution. Each node in the net is a very simple processor containing words which performs a few elementary tasks to alter the words, send and receive them to/from other processors. Despite the simplicity of each processor, the entire net can carry out very complex tasks efficiently. Many different works demonstrate the computational completeness of NEPs [4][10] and their ability to solve NP problems with linear or polynomial resources [11][2]. The emergence of such a computational power from very simple units acting in parallel is one of the main interests of NEPs.

NEPs can be used to accept families of languages. When they are used in this way they are called Accepting NEPs (ANEPs). Several variants of NEPs have been proposed in the scientific literature. NEP (the original model) [2], hibrid nets of evolutionary processor (HNEP) [4] and nets of splicing processors NEPS or NSP [10]. This last model uses a splicing processor, which adds a new operation (splicing rules) to mimic crossover in genetic systems. In section 3.1 we show an example of ANSP (the accepting variant of NSPs) solving the SAT problem. Nevertheless, all of them share the same general characteristics.

A NEP is built from the following elements: a) a set of symbols which constitutes the alphabet of the words which are manipulated by the processors, b) a set of processors, c) an underlying graph where each vertex represents a processor and the edges determine which processors are connected so they can exchange words, d) an initial configuration defining which words are in each processor at the beginning of the computation and e) one or more stopping rules to halt the NEP.

An evolutionary processor has three main components: a) a set of evolutionary rules to modify its words, b) some input filters that specifies which words can be received from other processors and c) an output filter that delimits which words can leave the processor to be sent to others. The variants of NEPs mainly differ in their evolutionary rules and filters. They perform very simple operations, like altering the words by replacing all the occurrences of a symbol by another, or filtering those words whose alphabet is included in a given set of words.

NEP's computation alternates evolutionary and communication steps: an evolutionary step is always followed by a communication step and vice versa. Computation follows the following scheme: when the computation starts, every processor has a set of initial words. At first, an evolutionary step is performed: the rules in each processor modify the words in the same processor. Next, a communication step forces some words to leave their processors and also forces the processors to receive words from the net. The communication step depends on the constraints imposed by the connections and the output and input filters. The model assumes that an arbitrary number of copies of each word exists in the processors, therefore all the rules applicable to a word are actually applied, resulting in a new word for each rule. The NEP stops when one of the stopping conditions is met, for example, when the set of

words in a specific processor (the ouput node of the net) is not empty. A detailed formal description of NEPs can be found in [1], [4] or [10].

1.2 Clusters of computers

Running NEPs simulators on cluster is one of the possible ways of explointing the inherent parallel nature of NEPs. The Java Virtual Machine (JVM), which can be considered the standard Java, cannot be run on clusters.

Several attempts have tried to overcome this limitation, for example: Java-Enabled Single-System-Image Computing Architecture 2 (JESSICA2) [8], the cluster virtual machine for Java developed by IBM (IBM cJVM) [3], Proactive PDC [12], DO! [9], JavaParty [6], and Jcluster [7].

The simulator described in this paper has been developed with both JVM and JavaParty.

2 jNEP

A lot of research effort has been devoted to the definition of different families of NEPs and to the study of their formal properties, such as their computational completness and their ability to solve NP problems with polynomial performance. However, no relevant effort, apart from [5], has tried to develop a NEP simulator or any kind of implementation. Unfortunately, the software described in this reference gives the possibility of using only one kind of rules and filters and, what is more important, violates two of the main principles of the model: 1) NEP's computation should not be deterministic and 2) evolutionary and communication steps should alternate strictly. Indeed, the software is focused in solving decision problems in a parallel way, rather than simulating the NEP model with all its details.

jNEP tries to fill this gap in the literature. It is a program written in Java which is capable of simulating almost any NEP in the literature. In order to be a valuable tool for the scientific community, it has been developed under the following principles: a) it rigorously complies with the formal definitions found in the literature; b) it serves as a general tool, by allowing the use of the different NEP variants and is ready to adapt to future possible variants, as the research in the area advances; c) it exploits as much as possible the inherent parallel/distributed nature of NEPs.

The jNEP code is freely available in http://jnep.e-delrosal.net.

2.1 jNEP design

jNEP offers an implementation of NEPs as general, flexible and rigorous as has been described in the previous paragraphs. As shown in figure 1, the design of the NEP class mimics the NEP model definition. In jNEP, a NEP is composed of evolutionary processors and an underlying graph (attribute *edges*) to define the net topology and the allowed inter processor interactions. The *NEP* class coordinates the main dynamic of the computation and rules the processors (instances of the *EvolutionaryProcessor* class), forcing them to perform alternate evolutionary and communication steps. It also stops the computation when needed. The core of the model includes these two classes, together with the *Word* class, which handles the manipulation of words and their symbols.

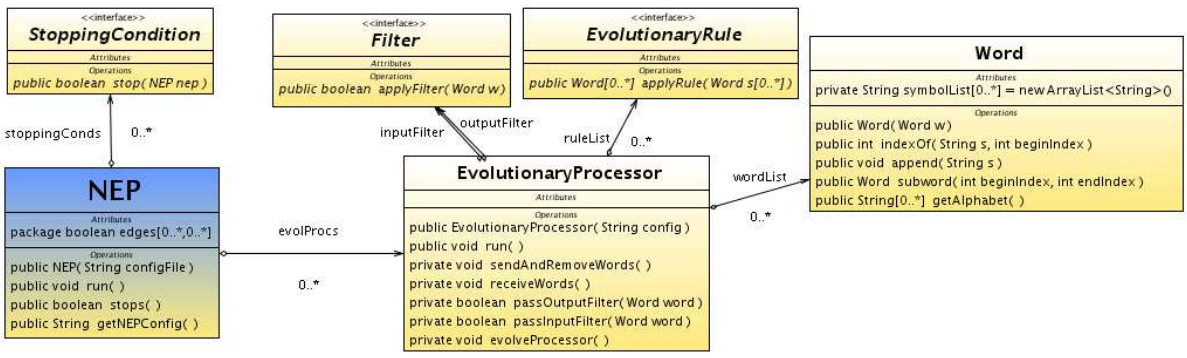


Figure 1: Simplified class diagram of jNEP

We keep *jNEP* as general and rigorous as possible by means of the following mechanisms: Java interfaces and the develop of different versions to widely exploit the parallelism availble in the hardware platform.

jNEP offers three interfaces: a) *StoppingCondition*, which provides the method *stop* to determine whether a *NEP* object should stop according to its state; b) *Filter*, whose method *applyFilter* determines which objects of class *Word* can pass it and c) *EvolutionaryRule*, which *appl*ies a *Rule* to a set of *Word*s to get a new set. *jNEP* tries to implement a wide set of NEPs' features. The *jNEP user guide* (http://jnep.e-delrosal.net) contains the updated list of filters, evolutionary rules and stopping conditions implemented.

Currently *jNEP* has two list of choices to select the parallel/distributed platform on which it runs (any combination of them is also available in http://jnep.e-delrosal.net). Concurrency is implemented by means of two different Java approaches: *Thread*s and *Process*es. The first needs more complex synchronization mechanisms. The second uses heavier concurrent threads. The supported platforms are standard JVM and clusters of computers (by means of JavaParty).

3 jNEP in practice

jNEP is written in Java, therefore to run jNEP one needs a Java virtual machine (version 1.4.2 or above) installed in a computer. Then one has to write a configuration file describing the NEP. The *jNEP user guide* (available at http://jnep.e-delrosal.net) contains the details concerning the commands and requirements needed to launch jNEP. In this section, we want to focus on the configuration file which has to be written before running the program, since it has some complex aspects important to be aware of the potentials and possibilities of jNEP.

The configuration file is an XML file specifying all the features of the NEP. Its syntax is described below in BNF format, together with a few explanations. Since BNF grammars are not capable of expressing context-dependent aspects, context-dependent features are not described here. Most of them have been explained informally in the previous sections. Note that the traditional characters <> used to identify non-terminals in BNF have been replaced by [ ] to avoid confusion with the use of the <> characters in the XML format.

- [configFile] ::= <?xml version="1.0"?> <NEP nodes="[integer]"> [alphabetTag] [graphTag] [processorsTag] [stoppingConditionsTag] </NEP>
- [alphabetTag] ::= <ALPHABET symbols="[symbolList]"/>
- [graphTag] ::= <GRAPH> [edge] </GRAPH>
- [edge] ::= <EDGE vertex1="[integer]" vertex2="[integer]"/> [edge]
- [edge] ::= λ
- [processorsTag] ::= <EVOLUTIONARY_PROCESSORS> [nodeTag] </EVOLUTIONARY_PROCESSORS>

The above rules show the main structure of the NEP: the alphabet, the graph (specified through its vertices) and the processors. It is worth remembering that each processor is identified implicitly by its position in the processors tag (first one is number 0, second is number 1, and so on).

- [stoppingConditionsTag] ::= <STOPPING_CONDITION> [conditionTag] </STOPPING_CONDITION>
- [conditionTag] ::= <CONDITION type="MaximumStepsStoppingCondition" maximum="[integer]"/> [conditionTag]
- [conditionTag] ::= <CONDITION type="WordsDisappearStoppingCondition" words="[wordList]"/> [conditionTag]
- [conditionTag] ::= <CONDITION type="ConsecutiveConfigStoppingCondition"/> [conditionTag]
- [conditionTag] ::= <CONDITION type="NonEmptyNodeStoppingCondition" nodeID="[integer]"/> [conditionTag]
- [conditionTag] ::= λ

The syntax of the stopping conditions shows that a NEP can have several stopping conditions. The first one which is met causes the NEP to stop. The different types try to cover most of the stopping conditions used in the literature. If needed, more of them can be added to the system easily. The *jNEP user guide* explains their semantics in detail.

- [nodeTag] ::= <NODE initCond="[wordList]" [auxWordList]> [evolutionaryRulesTag] [nodeFiltersTag] </NODE> [nodeTag]
- [nodeTag] ::= λ
- [auxWordList] ::= λ | auxiliaryWords="[wordList]"
- [evolutionaryRulesTag] ::= <EVOLUTIONARY_RULES> [ruleTag] </EVOLUTIONARY_RULES>
- [ruleTag] ::= <RULE ruleType="[ruleType]" actionType="[actionType]" symbol="[symbol]" newSymbol="[symbol]"/> [ruleTag]
- [ruleTag] ::= <RULE ruleType="splicing" wordX="[symbolList]" wordY="[symbolList]" wordU="[symbolList]" wordV="[symbolList]"/> [ruleTag]
- [ruleTag] ::= <RULE ruleType="splicingChoudhary" wordX="[symbolList]" wordY="[symbolList]" wordU="[symbolList]" wordV="[symbolList]"/> [ruleTag]

- [ruleTag] ::= λ
- [ruleType] ::= insertion | deletion | substitution
- [actionType] ::= LEFT | RIGHT | ANY
- [nodeFiltersTag] ::= [inputFilterTag] [outputFilterTag]
- [nodeFiltersTag] ::= [inputFilterTag]
- [nodeFiltersTag] ::= [outputFilterTag]
- [nodeFiltersTag] ::= λ
- [inputFilterTag] ::= <INPUT [filterSpec]/>
- [outputFilterTag] ::= <OUTPUT [filterSpec]/>
- [filterSpec] ::= type=[filterType] permittingContext="[symbolList]" forbiddingContext="[symbolList]"
- [filterSpec] ::= type="SetMembershipFilter" wordSet="[wordList]"
- [filterSpec] ::= type="RegularLangMembershipFilter" regularExpression="[regExpression]"
- [filterType] ::= 1 | 2 | 3 | 4

The preceding set of rules describe the elements of the processors: their initial conditions, rules, and filters. We have applied the same philosophy as in the case of stopping conditions, which means that our systems supports almost all kinds found in the literature at the moment. Future types can also be added. The reader may refer to the *jNEP user guide* for further detailed information.

- [wordList] ::= [symbolList] [wordList]
- [wordList] ::= λ
- [symbolList] ::= `a string of symbols separated by the character '\_'`
- [boolean] ::= true | false
- [integer] ::= `an integer number`
- [regExpression] ::= `a Java regular expression`

3.1 An example: solving the SAP problem with linear resources

Reference [10] describes a NEP with splicing rules (ANSP) which solves the boolean satisfiability problem (SAT) with linear resources, in terms of the complexity classes also present in [10]. We can use jNEP to actually build and run this ANSP. The following is a broad summary of the config file for such a ANSP, applied to the solution of the SAT problem for three variables. The entire file can be downloaded from jnep.e-delrosal.net.

```
<NEP nodes="9">
  <ALPHABET symbols="A_B_C_!A_!B_!C_AND_OR_(_)_[A=1]_[B=1]_[C=1]_[A=0]_[B=0]_[C=0]_#_UP_{_}_1"/>
  <!-- WE IGNORE THE GRAPH TAG TO SAVE SPACE. THIS NEP HAVE A COMPLETE GRAPH -->
  <STOPPING_CONDITION>
    <CONDITION type="NonEmptyNodeStoppingCondition" nodeID="1"/>
  </STOPPING_CONDITION>
  <EVOLUTIONARY_PROCESSORS>
    <NODE initCond="{_(_A_)_AND_(_B_OR_C_)_}" auxiliaryWords="{_[A=1]_# {_[A=0]_# {_[B=1]_#
                                                              {_[B=0]_# {_[C=1]_# {_[C=0]_#">  <!-- INPUT NODE -->
      <EVOLUTIONARY_RULES>
        <RULE ruleType="splicing" wordX="{" wordY="(" wordU="{_[A=1]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="(" wordU="{_[A=0]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[A=0]" wordU="{_[B=0]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[A=0]" wordU="{_[B=1]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[A=1]" wordU="{_[B=0]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[A=1]" wordU="{_[B=1]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[B=0]" wordU="{_[C=0]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[B=0]" wordU="{_[C=1]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[B=1]" wordU="{_[C=0]" wordV="#"/>
        <RULE ruleType="splicing" wordX="{" wordY="[B=1]" wordU="{_[C=1]" wordV="#"/>
      </EVOLUTIONARY_RULES>
      <FILTERS>
        <INPUT type="4" permittingContext="" forbiddingContext="[A=1]_[B=1]_[C=1]_[A=0]_[B=0]_[C=0]_#_UP_{_}_1"/>
        <OUTPUT type="4" permittingContext="[C=1]_[C=0]" forbiddingContext=""/>
      </FILTERS>
    </NODE>
    <NODE initCond="">  <!-- OUTPUT NODE -->
      <EVOLUTIONARY_RULES>
      </EVOLUTIONARY_RULES>
      <FILTERS>
        <INPUT type="1" permittingContext="" forbiddingContext="A_B_C_!A_!B_!C_AND_OR_(_)"/>
        <OUTPUT type="1" permittingContext="" forbiddingContext="[A=1]_[B=1]_[C=1]_[A=0]_[B=0]_[C=0]_#_UP_{_}_1"/>
      </FILTERS>
    </NODE>
```

```
    <NODE initCond="" auxiliaryWords="#_[A=0]_} #_[A=1]_} #_} #_1_)_}"> <!-- COMP NODE -->
      <EVOLUTIONARY_RULES>
        <RULE ruleType="splicing" wordX="" wordY="A_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="!A_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="B_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="!B_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="C_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="!C_OR_1_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="AND_(_1_)_}" wordU="#" wordV="}"/>
        <RULE ruleType="splicing" wordX="" wordY="[A=1]_(_1_)_}" wordU="#" wordV="[A=1]_}"/>
        <RULE ruleType="splicing" wordX="" wordY="[A=0]_(_1_)_}" wordU="#" wordV="[A=0]_}"/>
      </EVOLUTIONARY_RULES>
      <FILTERS>
        <INPUT type="1" permittingContext="1" forbiddingContext=""/>
        <OUTPUT type="1" permittingContext="" forbiddingContext="#_1"/>
      </FILTERS>
    </NODE>
    <NODE initCond=""  auxiliaryWords="#_1_)_} #_)_}">  <!-- A=1 NODE -->
      <EVOLUTIONARY_RULES>
        <RULE ruleType="splicing" wordX="" wordY="A_)_}" wordU="#" wordV="1_)_}"/>
        <RULE ruleType="splicing" wordX="" wordY="(_!A_)_}" wordU="#" wordV="UP"/>
        <RULE ruleType="splicing" wordX="" wordY="OR_!A_)_}" wordU="#" wordV=")_}"/>
        <RULE ruleType="splicing" wordX="" wordY="B_)_}" wordU="#" wordV="UP"/>
        <RULE ruleType="splicing" wordX="" wordY="C_)_}" wordU="#" wordV="UP"/>
      </EVOLUTIONARY_RULES>
      <FILTERS>
        <INPUT type="1" permittingContext="[A=1]" forbiddingContext="[A=0]_1"/>
        <OUTPUT type="1" permittingContext="" forbiddingContext="#_UP"/>
      </FILTERS>
    </NODE>
    <NODE initCond=""  auxiliaryWords="#_1_)_} #_)_}">  <!-- A=0 NODE -->
      <EVOLUTIONARY_RULES>
        <RULE ruleType="splicing" wordX="" wordY="OR_A_)_}" wordU="#" wordV=")_}"/>
        <RULE ruleType="splicing" wordX="" wordY="(_A_)_}" wordU="#" wordV="UP"/>
        <RULE ruleType="splicing" wordX="" wordY="!A_)_}" wordU="#" wordV="1"/>
        <RULE ruleType="splicing" wordX="" wordY="B_)_}" wordU="#" wordV="UP"/>
        <RULE ruleType="splicing" wordX="" wordY="C_)_}" wordU="#" wordV="UP"/>
      </EVOLUTIONARY_RULES>
      <FILTERS>
        <INPUT type="1" permittingContext="[A=0]" forbiddingContext="[A=1]_1"/>
        <OUTPUT type="1" permittingContext="" forbiddingContext="#_UP"/>
      </FILTERS>
    </NODE>
    <!-- NODES FOR 'B' AND 'C' ARE ANALOGOUS TO THOSE FOR 'A'. WE DO NOT PRESENT THEM TO SAVE SPACE-->
  </EVOLUTIONARY_PROCESSORS>
</NEP>
```

With this config file, at the end of its computation, jNEP outputs the interpretation which satisfies the logical formula contained in the file, namely:

```
(_A_)_AND_(_B_OR_C_):{_[C=0]_[B=1]_[A=1]_} {_[C=1]_[B=1]_[A=1]_} {_[C=1]_[B=0]_[A=1]_}
```

This ANSP is able to solve any formula with three variables. The formula to be solved must be specified as the value of the *initCond* attribute for the input node.

4 Conclusions and further research lines

jNEP is one of the first and more complete implementations of the family of abstract computing devices called NEPs. *jNEP* simulates not only the basic model, but also some of its variants, and is able to run on clusters of computers.

In the future we plan to offer full acces to the cluster version by means of the web. We also plan to develop a graphic user interface to ease the definition of the NEP being simulated. *jNEP* will be used as a module in the design of an automatic programming methodology to design NEPs to solve a given problem.

Acknowledgement: This work was supported in part by the Spanish Ministry of Education and Science (MEC) under Project TSI2005-08225-C07-06.

References

[1] J. Castellanos, C. Martin-Vide, V. Mitrana, and J. M. Sempere. "Networks of evolutionary processors". *Acta Informatica*, 39(6-7):517-529, 2003.

[2] Juan Castellanos, Carlos Martin-Vide, Victor Mitrana, and Jose M. Sempere. “Solving NP-Complete Problems With Networks of Evolutionary Processors.” *Connectionist Models of Neurons, Learning Processes and Artificial Intelligence : 6th International Work-Conference on Artificial and Natural Neural Networks*, IWANN 2001 Granada, Spain, June 13-15, Proceedings, Part I, 2001.

[3] http://www.haifa.il.ibm.com/projects/systems/cjvm/index.html

[4] E. Csuhaj-Varju, C. Martin-Vide, and V. Mitrana. “Hybrid networks of evolutionary processors are computationally complete.” *Acta Informatica*, 41(4-5):257-272, 2005.

[5] M. A. Diaz, N. Gomez Blas, E. Santos Menendez, R. Gonzalo, and F. Gisbert. “Networks of evolutionary processors (nep) as decision support systems.” In Fith International Conference. *Information Research and Applications*, volume 1, pages 192-203. ETHIA, 2007.

[6] http://wwwipd.ira.uka.de/JavaParty/

[7] http://vip.6to23.com/jcluster/

[8] http://i.cs.hku.hk/ wzzhu/jessica2/index.php

[9] Pascale Launay, Jean-Louis Pazat. “A Framework for Parallel Programming in Java.” *INRIA Rapport de Recherche* Publication Internet - 1154 decembre 1997 - 13 pages

[10] Florin Manea, Carlos Martin-Vide, and Victor Mitrana. “Accepting networks of splicing processors: Complexity results.” *Theoretical Computer Science*, 371(1-2):72-82, February 2007.

[11] Florin Manea, Carlos Martin-Vide, and Victor Mitrana. “All np-problems can be solved in polynomial time by accepting networks of splicing processors of constant size.” *DNA Computing*, pages 47-57, 2006.

[12] http://www-sop.inria.fr/sloop/javall/

Emilio del Rosal, Rafael Nuñez, Carlos Castañeda, and Alfonso Ortega
Universidad Autónoma de Madrid
Departamento de Ingeniería Informática
Av/Francisco Tomás y Valiente 7 - 28049 Madrid (Spain)
E-mail: emilio.delrosal@uam.es

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 486-490

Reducing the Number of Processors Elements in Systolic Arrays for Matrix Multiplication using Linear Transformation Matrix

Halil Snopce, Lavdrim Elmazi

Abstract:
Besides different definitions, in this work is given the so called transformation matrix, which maps the given index space in another index space. Transformation used in this new index space reduces the number of processing elements in the array. We illustrate all possible instances of transformation matrices and we show the importance of using the transformation matrix by comparing the number of processing elements of the array where we use it with another array where this transformation is not used. For this purpose also is given a mathematical explanation. The comparison is made using the matrices of size N=4.
Keywords: Systolic array, matrix multiplication, linear transformation matrix, number of processor elements.

1 Introduction

Matrix multiplication plays a crucial role in many scientific disciplines. This multiplication can be thought of as the main tool for many other computations in different areas, like those in seismic analysis, different simulations (like galactic simulations), aerodynamic computations, signal and images processing etc. In this paper is using a special design named systolic arrays which are suitable for matrix multiplication algorithm and offer both pipelineability and parallelism. On the area of systolic designs there are two main questions: the first one is how to choose the appropriate systolic array for certain application and the second question is how to minimize the number of processors. The main result in this work gives a possible answer for the second question mentioned above

2 Definition of systolic arrays

Definition 1. (Rao-[11])- A systolic array is a network of processors in which the processors can be placed at the grid points of a finite lattice so that:

- Topologically: If there is directed link from the processor at location I to the processor at location $(I+d)$ for some d, then there is such a link for every I within the lattice.
- Computationally: If a processor receives a value on an input link at time t, then it receives a value at time $(t+\Delta)$ on the corresponding output link, where Δ is time period that is independent of the network size, the orientation of the link and the location of the processor.

Definition 2. (Kung-[12])-A systolic array is a computing network possessing with the features of: synchrony, modularity and regularity, spatial locality, pipelinability, repeatability and high parallelism.

3 Systolic array for matrix multiplication using linear transformation matrix

Let A and B be two matrices of size NxN and we consider the problem of finding the resulting matrix C using the algorithm for matrix multiplication given below:

Algorithm 1

for $i,j,k = 1$ *to* N
$a(i,j,k) = a(i,j-1,k)$; $b(i,j,k) = b(i-1,j,k)$; $c(i,j,k) = c(i,j,k-1) + a(i,j,k-1) \cdot b(i,j,k-1)$
end
where $a(i,0,k) = a_{ik}$; $b(0,j,k) = b_{kj}$; $c(i,j,0) = 0$

Let $P_{ind} = \{(i,j,k)/1 \le i,j,k \le N\}$ be index space of used and computed data for matrix multiplication. Then we define the linear transformation matrix T given below:

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \tag{1}$$

Where $T_1 = [t_{11}\ t_{12}\ t_{13}]$ is the scheduling vector (for matrix multiplication is [1 1 1])and $S = [T_2\ T_3]^T$ is transformation which maps P_{ind} into 2-dimensional systolic array. Data dependency matrix for $Algorithm$ 1 is given with:

$$D = \begin{pmatrix} \overrightarrow{e}^3_b & \overrightarrow{e}^3_a & \overrightarrow{e}^3_c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The matrix T is associated with the so called projection direction $u = [u_1\ u_2\ u_3]^T$ (there are some possible allowable projection vectors, see [1]), so that the following conditions must satisfied:

$$detT \neq 0 \tag{2}$$

$$T_2 u = 0 \ and \ T_3 u = 0 \tag{3}$$

$$\Delta_S = SD \qquad \text{have to be from the set } \{-1,0,1\} \tag{4}$$

The transformation matrix T maps the index point $(i,j,k) \in P_{ind}$ into the point $(t,x,y) \in T \cdot P_{ind}$ where: $t = T_1[i\ j\ k]^T = [1\ 1\ 1][i\ j\ k]^T = i+j+k$ and

$$[x\ y]^T = S[i\ j\ k]^T \qquad \text{for} \quad (i,\ j,\ k) \in P_{ind} \tag{5}$$

t is time where calculations are performed, and (x,y) are the coordinates of processors elements on 2-dimensional systolic array. If $P_{in} = \{P_{in}(a), P_{in}(b), P_{in}(c)\}$ is space of initial computations with: $P_{in}(a) = \{(i,0,k)/\ 1 \le i,k \le N\}$; $P_{in}(b) = \{(0,j,k)/\ 1 \le i,j \le N\}$; $P_{in}(c) = \{(i,j,0)/\ 1 \le i,j \le N\}$ Then for the new position of the vector γ, $\gamma \in \{a,b,c,\}$ is taken: $p^*_\gamma = p_\gamma - (i+j+k-2)e^3_\gamma \Rightarrow [x\ y]^T_\gamma = S \cdot p^*_\gamma$. Let us consider the case where $u = [1\ 1\ 1]^T$. From (3) we have:

$$t_{21}+t_{22}+t_{23} = 0 \text{ and } t_{31}+t_{32}+t_{33} = 0 \tag{6}$$

Considering (1), (2), (4) and (6), below are given all possible transformation matrices:

$$\begin{pmatrix} 1 & 1 & 1 \\ \pm 1 & 0 & \mp 1 \\ \mp 1 & \pm 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ \pm 1 & \mp 1 & 0 \\ \pm 1 & 0 & \mp 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ \mp 1 & \pm 1 & 0 \\ 0 & \mp 1 & \pm 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 0 & \pm 1 & \mp 1 \\ \mp 1 & \pm 1 & 0 \end{pmatrix},$$

To implement the mapping $T : (i,j,k) \longrightarrow (t,x,y)$, first we define a linear mapping $L = (L_1, L_2)$ such that: $T \circ L :\ P_{ind} \longrightarrow \overline{P}_{ind}$ where $L :\ P_{ind} \longrightarrow P^*_{ind}$ and $T :\ P^*_{ind} \longrightarrow \overline{P}_{ind}$.

Definition 3. For $Algorithm$ 1, with the projection direction $u = [1\ 1\ 1]^T$ the mapping $L = (L_1, L_2)$ is defined in two possible cases, given below:

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, L_2 = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} \ or \ L_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, L_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} \tag{7}$$

The elements $(u,v,w) \in P^*_{ind}$ will be obtained from: $[u\ v\ w]^T = L_1[i\ j\ k]^T + L_2$. Let transformation matrix be:

$$T = \begin{pmatrix} T_1 \\ S \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

If we take matrix L given with (7) (first form) we have:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = L_1 \begin{pmatrix} i \\ j \\ k \end{pmatrix} + L_2 = \begin{pmatrix} i \\ i+j-1 \\ i+k-1 \end{pmatrix}$$

From (5) for the new vector (u,v,w) the position of PE is $[x\ y]^T = [1-j,\ 1-k]^T$ and the new initial space will be: $\hat{P}_{in}(a) = \{(i,0,i+k-1)/\ 1 \le i,k \le N\}$; $\hat{P}_{in}(b) = \{(0,i+j-1,i+k-1)/\ 1 \le i,j,k \le N\}$; *and* $\hat{P}_{in}(c) = \{(i,i+j-1,0)/\ 1 \le i,j \le N\}$. Then we have: $P^*_{in}(a) = [i,\ 0,\ i+k-1]^T - (i+0+i+k-1-3+1)\cdot[0\ 1\ 0]^T = [i,\ 3-2i-k,\ i+k-1]^T$; $P^*_{in}(b) = [4-2i-j-k,\ i+j-1,\ i+k-1]^T$ *and* $P^*_{in}(c) = [i,\ i+j-1,\ 3-2i-j]^T$. Now we can find the positions of input data in the array: $a(i,0,i+k-1) \longrightarrow [x\ y]^T_a = S\cdot P^*_{in}(a) = [3i+k-3,\ 1-k]^T$; $b(0,i+j-1,i+k-1) \longrightarrow [5-3i-2j-k,\ 5-3i-j-2k]^T$ and $c(i,i+j-1,0) \longrightarrow [1-j,\ 3i+j-3]^T$. Communication links are given with:

$$\Delta_S = S\cdot D = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \overrightarrow{e}^2_b & \overrightarrow{e}^2_a & \overrightarrow{e}^2_c \end{pmatrix}$$

So, the corresponding systolic array for $N = 4$ using mapping L is given in fig.1.

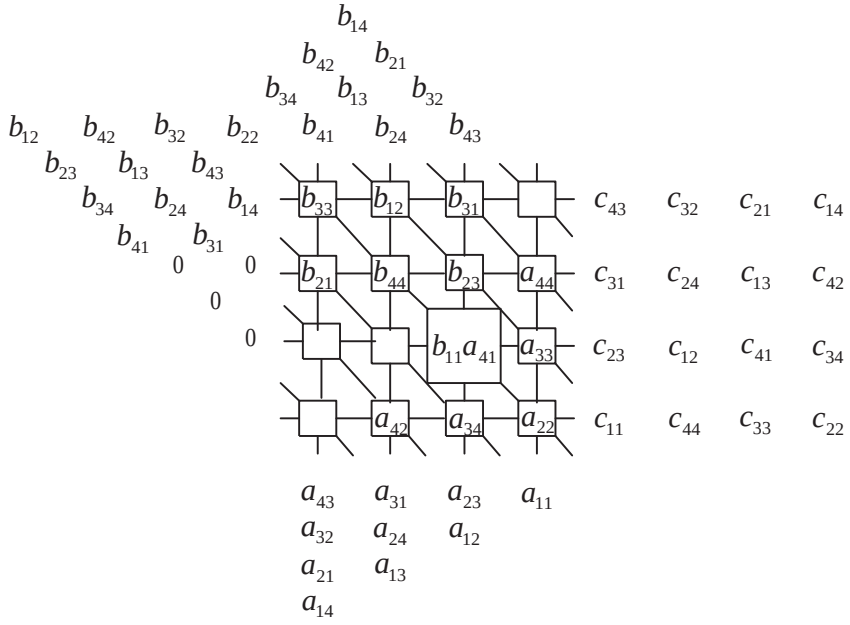


Figure 1: systolic array for $N = 4$ using mapping L

4 Systolic array for matrix multiplication without using of linear transformation matrix

On the other hand, if we do not use transformation L, and if we take the transformation matrix given with:

$$T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \tag{8}$$

Then we will have: $[x\ y]^T = S[i\ j\ k]^T = [i-k,\ j-k]^T$, $1 \le i,j,k \le N$ and $P^*_a = [i,\ 2-i-k,\ k]^T$; $P^*_b = [2-j-k,\ j,\ k]^T$ *and* $P^*_c = [i,\ j,\ 2-i-j]^T$. Now we can find the positions of input data in the array: $a(i,0,k) \longrightarrow [x\ y]^T_a = S\cdot P^*_a = [i-k,\ 2-i-2k]^T$; $b(0,j,k) \longrightarrow [2-j-2k,\ j-k]^T$ and $c(i,j,0) \longrightarrow [2i+j-2,\ 2j+i-2]^T$. Communication links are given with:

$$\Delta_S = S\cdot D = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} \overrightarrow{e}^2_b & \overrightarrow{e}^2_a & \overrightarrow{e}^2_c \end{pmatrix}$$

The corresponding systolic array for $N = 4$(known like SHSA array) without using mapping L is given in fig.2.

Without proof we will give the theorem (taken from [7]) for 3-nested loop algorithm which is valid and for *Algorithm* 1, taking on consideration that matrix multiplication is special case of 3-nested loop algorithm.

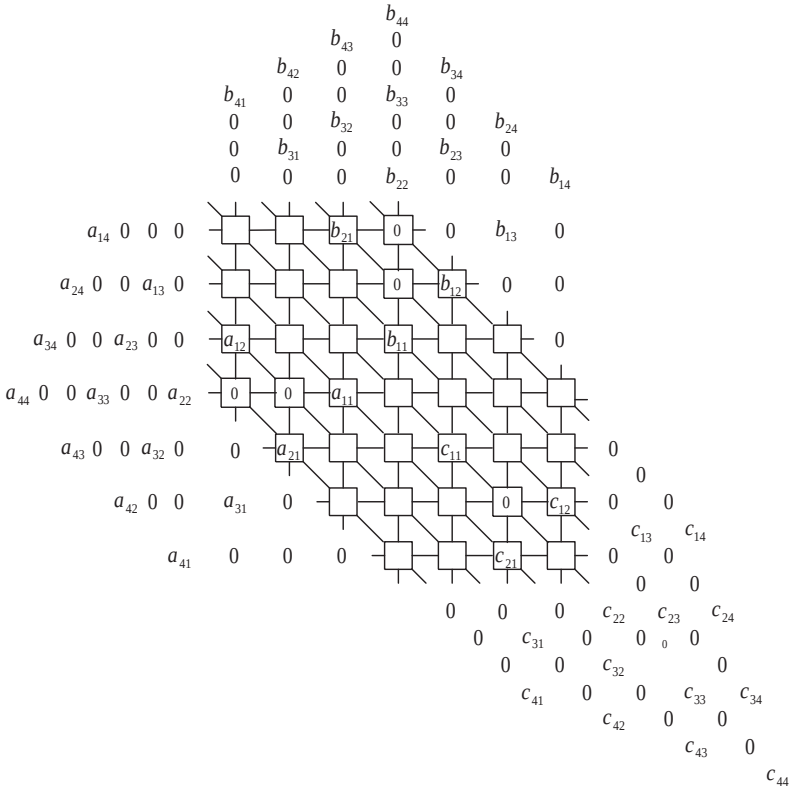


Figure 2: the SHSA array for $N = 4$ (without using mapping L)

Theorem 4. *([7]): The number of processors on the array for 3-nested loop algorithm with the size of loops (N_1, N_2, N_3) is given by:*

$$\Omega = \begin{cases} N_1N_2N_3 & \text{if } a_i > N_i \quad \text{for } 1 \le i \le 3 \\ N_1N_2N_3 - \omega, & \text{otherwise} \end{cases}$$

where: $\omega = (N_1 - a_1)(N_2 - a_2)(N_3 - a_3)$ and $a_i = \left|\frac{T_{1i}}{gcd(T_{11},T_{12},T_{13})}\right|$ where $T_{1i}, i = 1,2,3$ is $(1,i)$ cofactor (minor) of matrix T and $gcd(T_{11},T_{12},T_{13})$ is the greatest common divisor of T_{11}, T_{12} and T_{13}.

Using the above theorem, we have that the number of processors on SHSA array (which can be seen from Figure 2) is: $\Omega = N^3 - (N-1)^3 = N^3 - N^3 + 3N^2 - 3N + 1 = 3N^2 - 3N + 1$. In this case we have used that $N_1 = N_2 = N_3 = N$ and for the transformation matrix given with (8) we have: $|T_{11}| = |T_{12}| = |T_{13}| = 1$ from where $a_1 = a_2 = a_3 = 1$. For $N = 4$ we have that $\Omega = 37$.

But we observed that the number of processor elements where we used the mapping L is 16 (Figure 1). Below we give the theorem associated with the proof which confirms this:

Theorem 5. *The number of processing elements in 2-dimensional systolic array for the algorithm of matrix-matrix multiplication (Algorithm 1) for which is used the projection direction $u = [1\ 1\ 1]^T$, could be reduced and given with $\Omega = N^2$.*

Proof. We saw that for *Algorithm* 1 with $u = [1\ 1\ 1]^T$ could be applied the mapping L defined by (7). We have used transformation T given in (1). The composition $T \circ L$ was used for obtaining the corresponding systolic array, (because of the mapping $T \circ L : P_{ind} \longrightarrow \overline{P}_{ind}$). Therefore the number of processor elements Ω depends on the matrix $T \circ L$. But because in $L = (L_1, L_2)$, the part $L_2 = [0\ -1\ -1]^T$ contains no variables, we can conclude that Ω depends only on composition $H = T \circ L_1$. Therefore (also using (6)) we will have:

$$H = T \circ L_1 = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} t_{11}+t_{12}+t_{13} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & t_{32} & t_{33} \end{pmatrix}$$

Now we can see that cofactors $H_{12} = H_{13} = 0$. Using theorem 4 we have that $a_2 = a_3 = 0$ and $a_1 = 1$ $\left(a_1 = \left|\frac{H_{11}}{gcd(H_{11},0,0)}\right| = 1\right)$. Therefore we will have (taking $N_1 = N_2 = N_3 = N$): $\Omega = N_1N_2N_3 - (N_1 - 1)N_2N_3 = N^3 - N^3 + N^2 = N^2$. □

Because of theorem 5, and taking on consideration that in construction of the array given in Figure 1 we are using the transformation L given with (7), the number of processors (which can be seen from Figure 1) is $\Omega = 4^2 = 16$.

From this we can conclude about the advantage of using the transformation L, because such kind of transformation reduces the number of processors on the array.

5 Conclusions

In this paper we have used two types of linear transformation matrix for showing the result of our conclusion. We also have used two theorems to determine the number of processor elements on systolic arrays for matrix multiplication. We have emphasized the advantage of using the linear transformation L, which implicates exactly on the reduction of the number of processing elements. We have also gave models of discussed systolic arrays.

References

[1] M. P. Bekakos, "Highly Parallel Computations-Algorithms and Applications," *Democritus University of Thrace, Greece*, pp. 139-209, 2001.

[2] M.A. Frumkin , *Systolic Computations,* Scripps Research Institute, La Jolla, California, U.S.A., 1992.

[3] Esonu, M.O., Al-Khalili, A.J., Hariri, S. and Al-Khalili, D. , *Systolic Arrays: How to choose them,* IEE Proc., E-139, 3, pp. 179-188, 1992.

[4] Milentijevic, I.Z., Milovanovic, I.Z., E.I. and Stojcev, M.K. , *The Design of Optimal Planar Systolic Arrays for Matrix Multiplication,* Comput. Math. Appl., pp. 17-35, 1997

[5] Bekakos, M.P., Milovanovic, E.I., Milovanovic, I.Z. and Milentijevic, I.Z. , *An Efficient Systolic Array for Matrix Multiplication,* Proc. of the Fourth Hellenic European Conference on Computer Mathematics and its Applications (HERCMA '98), Athens '98, pp. 298-317, 1999

[6] Kung, H.T. and Leiserson, C.E. , *Systolic arrays for (VLSI), Introduction to VLSI Systems,* Addison-Wesley Ltd., Reading, MA, 1980.

[7] C.N. Zhang, J.H. Weston, Y. F. Yan. , *Determining object functions in systolic array designs,* IEEE Trans. VLSI Systems 2, No. 3 (1994), 357-360.

[8] Nam Ling and Magdy A Bayoumi. , *Specification and Verification of Systolic Arrays,* Santa Clara University and University of Southwestern Louisiana, 1999.

[9] Kung, H.T. and Leiserson, C.E. , *Algorithms for VLSI processor arrays,* Sparse Matrix Proceedings, SIAM Press, pp. 256-282, 1978.

[10] Brent, R.P., Kung, H.T., Luk, F.T. , *Some Linear-Time Algorithms for Systolic Arrays,* Cornell University Ithaca, NY, USA, pp. 863-876, 1983.

[11] S.K. Rao. , *Regular Iterative Algorithms and their implementation on Processor arrays* , Ph.D. Dissertation, Stanford University, Stanford, CA, 1985.

[12] S.Y.Kung. , *VLSI Array Processors* , Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[13] S. Ribaric. , *Arhitektura Racunala Pete Generacije* , Tehnicka Kniga-Zagreb, pp. 65-72, 1985

Halil Snopce, Lavdrim Elmazi
South East Europian University
Faculty of Computer Sciences and Technologies
Address: Ilindenska bb., 1200 Tetovo, Republic of Macedonia
E-mail: {h.snopce, l.elmazi }@seeu.edu.mk

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 491-496

Towards Great Challenge for Enterprise Science Within Knowledge-based Society Paradigm

Aurelian M. Stănescu, Lucian M. Ionescu, Adina Florea,
C. Şerbănescu, Mihnea A. Moisescu, Ioan S. Sacala

Abstract: The paper is aiming at launching the debate / forum to discuss the new scientific discipline of "e-Enterprise" (e-E) within the knowledge-based Society paradigm. There are great challenges to support the solid foundation of the new "science", copying with the requirements list of Knowledge-based society focusing on globalization and e-democracy approach for e-Enterprise. The first key-driver is the coherent and consistent new methodology for General System Theory based modeling, analysis, performance evaluation and conceptual design of e-Enterprise. The second one concerns with Digital World Theory and Applications based on advanced research in Mathematics, Physics and Computing Practice. A Romanian National Research funding-based project "REMEDIUM" (environment 4 health) is used as a complex study case.
Keywords: complex adaptive systems, metamodeling framework, digital world theory.

1 Introduction

The transition from the Postindustrial Society (PS) towards Knowledge-based Society(KbS) via The Information Society(IS) generates many paradigm shifts, as well as great challenges. All daily activities like healthcare systems, administration / e-government, education / lifelong learning/e-learning, culture business(from e-Commerce toward Knowledge - based Business), way of working/e-work and last, but not least Interoperability - focused networked Enterprise networking (virtual, extended, collaborative, a.s.o) are high priority research agenda of thousands of Labs worldwide.

One of the author's present paper (Aurelian M. Stanescu) has just contributed at key-issue, included in "Vision and Roadmap of Enterprise Interoperability" namely the "Systems and Complexity Science". Enterprises have many structures and relationships. Understanding their interactions is considered a major factor in contributing to the success of interoperability solutions and the performance of the network enterprise [Extended Enterprise (EE), Virtual Enterprise (VE), Collaborative Network Organization (CNO)]. The sustainable enterprises are COMPLEX ADAPTIVE SYSTEMS (CAS) [I. Dumitrache, A.M.Stanescu 2007]. Moreover, with reference to the concerted research roadmap for shaping business in the Knowledge-based Economy, they are also components within one or more innovative ecosystem [A.M.Stanescu 2007]. The evaluation of complexity in these "System of Systems" with "star-topology or" "pear - to - pear" ICT platform topology are in progress to be consolidated. Some researchers have attempted to extrapolate the results to a "General System Theory (GST)". This GTS could explain the behavior modeling of the CAS. This theory views all systems as dynamic, living entities that are goal-oriented and that evolve over time [Molina,2007]. Complexity science, generally considered as a branch of systems science, has been developed to address the emergence, adaptation, evolution, and self - organization of such systems. In particular, it concerns the coupling and interactions of the parts within these systems in a non linear fashion.

Mc Chutchy and F. Gensfeld, from Industrial System Division of M.I.T. has already highlighted the great challenge for not only modeling, but analysis, performance evaluation and synthesis of the CAS - concepts based socio - technological - economic systems, towards the contextual systems. In comparison with the "old" signal processing - based system (automatic control - oriented systems) the degree of complexity, as well as the degree of uncertainty are exponentially increased.

On the other hand, two new ideas has been created the context to develop our research topics. Prof. Shimon Nof, from the Purdue University/Prism Center provided two key - conferences [INCOM 06], [IFAC - MCPL]. These papers have just open the way to the holistic approach for the CAS involving key features of interoperability, scalability, self organizing, and availability vector.

The paper is concerned with launching the debate for a promising new scientific discipline "e - Enterprise" - Science. The structure of material is the following:

1. General System Theory - oriented Framework
2. Metamodeling-focused new approach for CAS in e-Enterprise

3. Digital World-Theory new coming science or provoking trend in co-sciences Mathematics, Physics and Computing Science

4. Conclusion and further work

2 General System Theory - oriented Framework

The companies are overheated due to the ICT- technologies. The well-known paradigm MONITOR/ANALYSE/PLAN/EXECUTE (MAPE) has been provided by the Department of Defense-USA (Air force Doctrine Center 1998) some time ago, but we also stress on another J.R. Boyd FEEDBACK PARADIGM O.O.D.A. (OBSERVE / ORIENT / DECIDE / ACT) addressing the decision-makers for every domain of real economy.

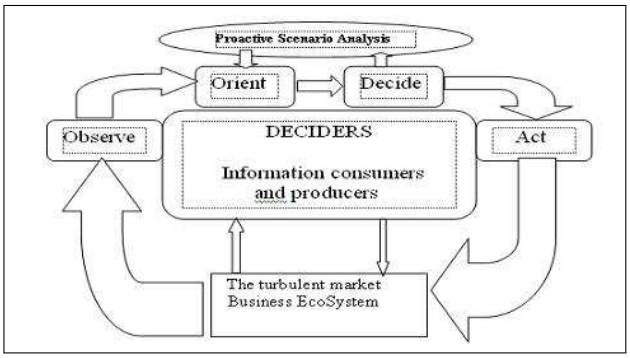


Figure 1: The OBSERVE-ORIENT-DECIDE-ACT (OODA) loop

An extension including ProActive Scenario Analysis is proposed at PROVE '08 IFIP Conference to be held in September 2008 [Stanescu 2008] as it is shown in fig.1.

Another paradigm is the "S.H.O.R. - STIMULATION-HYPOTHESIS-OPINION-RESPONSE" which is useful for the 'classical behaviorist' (psychology) to explicitly deal with uncertainties (Whol 1981).

In the following we must consider the meaning of these general features:

- Extending and enhancing scientific knowledge and truth about our existence.
- Using management of existing knowledge and truth about existence.
- Producing new technological knowledge through innovation.

The main paradigm the authors are concerned with in their present research is "Virtual Organization"/Collaborative Networks [Camarinha-Matos, Afsarmanesh, 2005] Taking into consideration the international research context, one could notice that the "concurrent engineering "(CE) paradigm, that has been developed rapidly since 1982, has to be re-balanced between advanced methodology useful in engineering science (including methods, tools, techniques- IST-CE.net.project phase 2 1999-2004- http://www.ce-net.org) into a collaborative science.

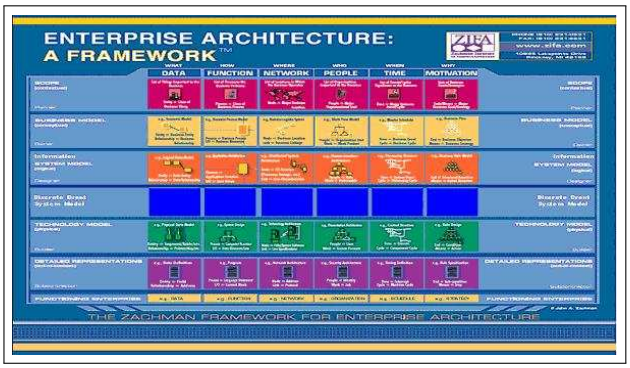


Figure 2: The Zachman Framework for Enterprise Architecture

The recent published document "Enterprise Interoperability: a concerted research roadmap for shaping business networking in the knowledge-based economy"- European Commission- "Information, society and media" ISBN-92-79-02437-x (http://publications.europa.en) coordinated by Man-Tze-Li, whilst one of the authors of the present paper (Aurelian Stanescu) was one of many contributors, provides the following list of indicating scientific disciplines, and whose ideas, propositions and findings could provide a starting point for the proposed science base.

1. Systems/Complexity Science
2. Information Science
3. Network Science
4. Web Science
5. Services Science
6. Economic Science
7. Social Science

The present paper is aiming at a scientific foundation of Distributed Concurrent (Collaborative) Engineering Science, which proves that the "best practices" era of the CE has just passed and our community is going to develop a more general systems theory-based foundation.

Recently, the public available document "Enterprise Interoperability: a concerted Research Roadmap for Shaping Business Networking in the knowledge-based Economy
(http://cordis.europa.eu/fp7/ict/enet/ei_en.html) coordinated by Man-Tze-Li provides the conceptual View of the ISU (Interoperability Service-Utility concept).

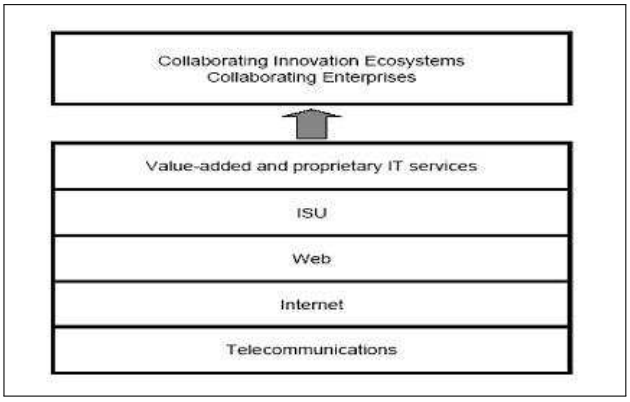


Figure 3: Conceptual View of the ISU.

The most important challenge raises by new scientific achievements on the Interoperability "problem-solutions". "Numerous knowledge links between various "actors" with a Collaborative Network enactment (enterprises, e-workers, a.s.o.) have to, rapidly and flexibly be combined to respond efficiently to market rapid changes. The size of an "enterprise" will matter for less than its ability to collaborate, its ability to adapt or its ability to interoperate" [Man-Tze-Li, 2006]. The most important challenge raised by new "trendy" research on operability. The problem is concerned with how to represent the complex System specifications addressing technical, semantic and pragmatic Interoperability [Dumitrache, 2007].

3 Metamodeling-focused new approach for CAS in e-Enterprise

Taking into consideration the last years advanced research in the domain of Adaptive Complex System used for consolidating e-Enterprise new coming science, one should highlight the following levels developing methodologies to analysis, performance evaluation and synthesis of e-Enterprise's metasystem.

- E terra modeling (C2)
- E meta modeling (C1)
- E modeling framework (C0)

1. Atomic Subsystems
2. API based Subsystems
3. Integrated Subsystems Platform
4. Functional / Behavioral e-Enterprise
5. Hypersystems
6. Metasystems: Collaborative Networked Organizations
7. Terra systems: Global e-Enterprise

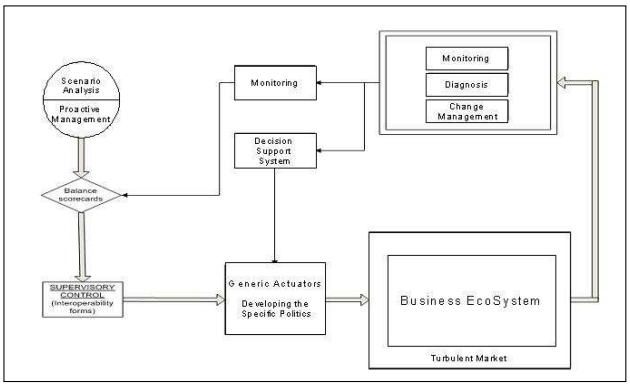


Figure 4: Modeling Framework levels, Systems of Systems hierarchy

Figure 5: CAS Metamodeling and Conceptual Design

To model and analyze the CAS e-Enterprise, a multiview approach will be focused on various views models as Functional Model, Informational Model, Organizational Model, Resource Model and Economic Model.

4 Digital World-Theory new coming science or provoking trend in co-sciences Mathematics, Physics and Computing Science

The present version of the DWT sketches the “object-oriented” Q++ language and provides implementation considerations.

Q++ is similar in purpose and presentation to any other computer programming language (automation / formal language / logic correspondence), except the gates process quantum digits.

The mathematics tools / technology needed to take into account quantum fluctuations, i.e. a dynamic hardware as opposed to a fixed hardware configuration is encapsulated into the QDR approach.

Its mathematical core the Homological Feynman Path Integral (Feynman processes on dg-coalgebra 2-categories) with its physical interface as Quantum Information Dynamics on Quantum Circuits (Non-commutative Hodge de Rham-Dirac Theory on the Quantum Dot Resolution), provides an "all-quantized-no-renormalization-needed" model of quantum information flow: random teleportation at the speed of light, the theoretical basis for non-conventional transportation methods.

There is no apriority space nor time or distance; we only model (eavesdropping) on a quantum communication through a quantum communication channel.

Our In-Out natural orientation makes us “see” only matter, most of the time. The random teleportation involves occasional random walk back in time we call “antimatter”; viewed in this way, it is retro causation!

The model interprets entropy as the generator of mass, which is a measure of symmetry. There is only one particle, the qubit, with dual functions: data and gate. The “rest” is achieved by the Internal-External duality which accounts for the missing energy and momentum from external space (“neutrino”), the 2:1 correspondence between the basic quark-pair (u,d) and lepton (e), with generations implemented later on as derived functions of the same 2:1 fundamental short exact sequence.

The doubling in the quantization process of current theories steams from the completion approach in the commutative world: fields of fractions (and then real numbers: Newton-Leibnitz space-time and analysis!). The “shortcut” is Hopf Theory!

Finally, quantum gravity is an organizational principle” organizing the quantum information sources and flow to minimize / optimize the “quantum thought” (Lorentz transformation as a 2-morphism to maximize / symmetrize the qubit flow).

In the literature, the universe is often compared with a quantum computer. May be it deserves more; it is a cognitive system, and we are just “half of the picture”. Observing and modeling our exterior coincides with its organization (cooling: lowering the entropy), not destined for a “thermal death”, and acquiring an entropic arrow for the “US-IT” permanent communication. Let’s not forget that we are not just “individuals”; we are connected in many ways we are only starting to explore...

Is the Digital World Theory “Science or Fiction”? It is a consistent scientific project in the design phase (not fully computational yet), with important conceptual and practical implications in the near future.

“Ultimate Particle Physics Theory” is Number Theory as the balance between the algebraic interpretation (Galois Theory) and geometric interpretation (Klein’s Program).

On the other hand (rather “hemisphere”) we should keep the real big picture: Reality is “The Quantum Matrix”!

“Everything” is quantum information flow, with its discrete quantum branching on rational Riemann surfaces as Pythagorean chords.

5 Conclusions and further research

The paper is aiming to launch an efficient debate on the subject: “Is necessary to provide more concepts in order to consolidate the Collaborative Network Organization as a solid holistic science?”

The conclusions are in this moment the following:

1. Yes, it is a must to consolidate the conceptual design of the CNO, but a General System Theory - oriented approach for a new methodology could be fertile.

2. The time of a “grand unification” has just arrived, but “the universe of discourse” is strong connected with Digital World Theory.

3. The main formal support should be the Discrete Event Dynamical Systems; the focus on intelligent supervisory control could be useful for every multidisciplinary research team.

4. One potential ambitious case study is going to be developed for co-domain Environment 2 Health CNO.

Further work is concerned with the development of the Oracle Data Center within the laboratory of Intelligent Information Systems. The work is in progress at University POLITEHNICA of Bucharest.

References

[1] Cassandras, La Fortune *Discrete Event Dynamic Systems* Klumwer Academic Publishers; pag. 31-52, 1999.

[2] INCOM 06 - A. Dolgui, X. Boucher , Information Systems, Control and Interoperability, Volume 1.

[3] I. Dumitrache, A. M. Stanescu, S. I. Caramihai *Complex Adaptive Systems* IFAC-MCPL Conference , ISBN, Sept 2007.

[4] H. Ehrig, W. Reisig, G. Rozenberg, H. Weber *Petri Net Technology for Communication-based Systems* Springer Verlaag ISBN 3-540-20538-1 pg 2-22.

[5] G. T. Horowich *String Theory without a Background spacetime Geometry-Mathematical Aspects of String Theory*, S. Yan Ed. , 127-140, 1987.

[6] L. M. Ionescu *Q++ and a Non-Standard Model* , Olimp Press, 2007.

[7] D. Karagiannis, H. Kuhn *C1-metamodelling oriented Business Process Monitoring and Management* [invited workshop in IFAC-INCOM 06] Conference - Proceedings EPSM Saint Ettiene.

[8] R. Kurzweil *Human 2.0* New Scientist, Review 2001.

[9] Cornelius Leondes *Intelligent knowledge-based Systems* Kluwer Academic Publishers (Vol.1), 2005.

[10] PROVE 07 Springer Verlag, Guimaraes, Sept. 2007.

[11] PROVE 05 Springer Verlag, Valencia, Sept. 2005.

[12] A. P. Sage *Special Issues for Human Decision Making and Command and Control*, *Information, Knowledge, Systems, Management* , IOS Press, N4, vol.2, 2001.

[13] O. Stanasila *Discrete Mathematics* , Tech Publisher, pg. 32-37, Bucharest 1985, 2007.

[14] A. M. Stanescu et all *e-Remedy environment 4 health* (2009 June RFID Call).

[15] A. M. Stanescu et all *From taxonomy towards ontology-based modelling framework in General System Theory*, paper accepted for ICCC Conference, Oradea, 15-17 May 2008.

[16] A. M. Stanescu et all *Textbook for M.Sc. sthudent* Discrete Event Dynamic Systems , Politehnica Press, 1999.

[17] A. Toffler: *The Third Wave*, Bonton Books, Inc, 19, pag 7-9.

[18] The 6th Framework programme (IST Priority) annual *Vision and Road Map* (http://cordis.europa.eu/ist/ict-ent-net/ei-roadmap_en.html), (Version 4:0 - 2006, version 5.6 - 2007, version 7 - 2008).

[19] www.athena-ip.org

[20] www.cordis.eu, Enterprise Interoperability

[21] www.ecolead.org

Aurelian M. Stănescu, Adina Florea
Mihnea Moisescu, Ioan-Ştefan Sacala
University POLITEHNICA of Bucharest
Automatic Control and Informatics Department
313 Splaiul Independentei, Sector 6, Bucharest,
Romania
E-mail: ams@cpru.pub.ro, amm@cpru.pub.ro
sacalaioan@yahoo.com

Lucian Ionescu
Illinois State University
Mathematics Department
Illinois, USA
E-mail: lmiones@ilsstu.edu

Cristina Şerbănescu
University POLITEHNICA of Bucharest
Mathematics Department
313 Splaiul Independentei, Sector 6, Bucharest,
Romania

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 497-502

Number of Efficient Points in some Multiobjective Combinatorial Optimization Problems

Milan Stanojević, Mirko Vujošević, Bogdana Stanojević

Abstract: The number of efficient points in criteria space of multiple objective combinatorial optimization problems is considered in this paper. The number of Pareto optimal solutions grows exponentially with the problem size. In this paper it is concluded that under certain assumptions, which are reasonable and applicable in the majority of practical problems, the number of efficient points grows polynomially.
Experimental results with the shortest path problem, the Steiner tree problem on graphs and the traveling salesman problem show that the number of efficient points is even much lower than the polynomial upper bound.
Keywords: multiobjective combinatorial optimization, Pareto optimal point, efficient point

1 Introduction

For all combinatorial problems cardinality of the feasible solution set grows exponentially with the problem size. For one group of combinatorial problems (e.g. the shortest path problem, the shortest spanning tree problem, assignment problem etc.) algorithms that can find a single-criterion problem solution in polynomial time are known. This problem class is denoted by $\mathcal{P}$. For other combinatorial problems (e.g. traveling salesman problem, Steiner tree problem, knapsack problem etc.) so called *non deterministic polynomial* algorithms exist. These problems belong to the class denoted by $\mathcal{NP}$. For this problem's class it is not proved whether polynomial algorithms exist or not. There is a third class of problems (e.g. finding all spanning trees for a given graph) for which it is known that they can be solved only by exponential algorithms. In such problems result usually consists of exponential amount of data, so the exponential time is needed just to represent them. Let's denote this class by $\mathcal{E}$.

In all the known literature which concerns multiobjective combinatorial optimization (MOCO), mostly the set of Pareto optimal solutions is being observed, and it has been stated that its size can grow exponentially with the problem size. Moreover, with sufficient number of uncorrelated criteria it is possible to achieve that every feasible solution is Pareto optimal [1]. This implies that there is no efficient method of determining all Pareto optimal solutions for problems of bigger dimensions, because such a procedure would not be even $\mathcal{NP}$ hard, but strictly exponential, i.e. it would belong to class $\mathcal{E}$. Such results are confirmed for many known problems, such as: the shortest spanning tree, the shortest path problem, traveling salesman problem, assignment problem and knapsack problem [2, 3, 4, 5, 6].

In Section 2 some multiobjective optimization models are introduced. Experimental results presented in Section 3 show that the number of efficient points is even much lower than the polynomial upper bound. In Section 4 some concluding remarks are formulated.

2 Multiobjective optimization models

The general model of multiobjective optimization problem can be briefly formulated as follows.

$$\min_{x \in X} f(x) \tag{1}$$

where x is the decision variable, X is the feasible solution set, and $f = (f_1, \ldots, f_p)$ is the p-dimensional vector of objective functions.

Connected to this model we will use the notions: Pareto optimal solution, efficient point in criteria space, Pareto set, efficient points set, marginal solution, ideal value, ideal point and nadir point, which are very well known in multiobjective optimization. All notions regarding Model (1) will be easily applied to the next models which will be defined as particular cases of it.

The model of multiobjective combinatorial optimization problem can be formulated as follows.

$$\min_{x \in \chi} f(x) \tag{2}$$

where x is the decision variable, χ is the feasible solution set of the model which is a subset of the power set of a finite set E (i.e. $\chi \subset \wp(E)$, $E = \{e_1, ..., e_m\}$), and $f = (f_1, ..., f_p)$ is the $p-$dimensional vector of objective functions.

Moreover, $x \in \chi$ can be represented by a binary m-dimensional variable $(x_1, x_2, ..., x_m) \in \{0,1\}^m$ where $x_k = 1$ if and only if the corresponding element $e_k \in E$ belongs to x and $x_k = 0$ if and only if the corresponding element $e_k \in E$ does not belong to x.

The experiments related to the number of efficient points of multiobjective optimization problems were done on three types of multiple objective network problems:

- Multiobjective shortest path problem (SPP),
- Multiobjective Steiner tree problem on graphs (STP) and
- Multiobjective traveling salesman problem (TSP).

For all the three problems, an undirected graph $G = (V, E)$ with $|V| = n$ nodes, $|E| = m$ edges and $w_k : E \to \mathbb{Z}^+$, $k = 1, 2, ..., p$ weight functions on the edges is given. In addition, for SPP starting node $s \in V$ and target node $t \in V$ and for STP set of terminal nodes $T \subset V$ are given.

For each of the problems, each feasible solution represents a specific graph structure. For SPP it is a path from s to t and the feasible set χ is a set of all such paths. For STP, χ represents set of all Steiner trees of graph G and terminal nodes T. And for TSP χ is a set of all Hamiltonian cycles.

For any of the problems, a feasible solution $x \in \chi$ is a set of edges that belong to a feasible graph structure. Alternatively, a feasible solution is a vector $(x_1, x_2, \ldots, x_m) \in \{0,1\}^m$ that satisfy a set of constraints specific to each problem.

The form of goal functions in each of above mentioned problems can be twofold, depending on the type of k-th criterion:

- When a criterion represents length or weight, the corresponding goal function is linear as follows

$$f_k(x) = \sum_{j=1}^{m} c_j^k x_j \tag{3}$$

  and it is minimized.

- When a criterion type represents capacity, the corresponding goal function is

$$f_k(x) = \min_{1 \le j \le m} c_j^k x_j \tag{4}$$

  and it is maximized.

Problems of minimizing a function of type (3) are called *minisum*, while problems of maximizing a function of type (4) are called *maximin* or *bottleneck* problems. Coefficients c_j^k, $j = 1, \ldots, m$, $k = 1, \ldots, p$ in general case are real numbers that represent length, height, weight or capacity of elements of set E. In our experiments it is presumed that coefficients c_j^k, $j = 1, \ldots, m$, $k = 1, \ldots, p$ have integer values. Practically that does not mean a loss of generality because in real life problems coefficients are rational numbers which can be transformed to integers.

In one multiobjective combinatorial optimization problem both kinds of goal function can exist. If functions of type (4) exist, they can be transformed to

$$f_k(x) = \max_{1 \le j \le m} \left(-c_j^k x_j \right), \tag{5}$$

that have to be minimized, in order to match the general formulation (2).

3 Experiments

Problem SPP, in its single criterion version, belongs to the class $\mathcal{P}$ and its multiobjective version represents one of the most studied problems of MOCO. The second problem (STP), in its *minisum* version, belongs to the class $\mathcal{NP}$, but in its *maximin* version belongs to the class $\mathcal{P}$ [8]. TSP, in both *minisum* and *maximin* versions is an $\mathcal{NP}$ hard problem.

3.1 Descriptions of the developed computer programs

For each of the problems, a specific computer program was developed by authors. Each program finds all efficient points and one solution for each of them. Also, it is announced if the efficient point is supported or non supported [1, 2]. For finding all efficient points, iterative ε-constraints method was applied. It is implemented only for two criteria problems.

The programs for SPP and STP support instances with both kinds of criteria *minisum* and *maximin*. Program for TSP supports only instances with *minisum* criteria type.

For all the experiments random instances with certain characteristics were generated. All instances had two non correlated criteria. Each result of experiments is obtained as an average of 10 randomly generated instances with the same characteristics.

For SPP problem instances had specific structure in order to provide paths to have at least $\sqrt{m}$ edges. Graph density varied between 36% and 60% depending on the number of vertexes. Instances with smaller number of vertexes had higher density.

Instances for examining STP problem were generated so each vertex has a certain probability to be connected to every other vertex. It was assure that all instances will not contain unconnected vertexes. Both, too dense or too spare graphs would not be proper for this kind of problem. Average density was 35% for graphs with 20 vertexes and 14% for graph with 50 vertexes, i.e. the average number of a vertex degree was 7 for both dimensions. The number of terminals was 5 for all instances.

TSP instances were generated in a similar way as STP, but density was higher, 50% for all graph dimensions.

3.2 Types of the performed experiments

Three groups of experiments were performed.

The first group was inspired by a supposition that the upper bound for the number of efficient points depends on the length of the intervals from which edges lengths can get integer values. Three such intervals were defined: I1=[0, 99], I2=[100, 999] and I3=[1000, 9999] (as sets of two, three and four digits numbers) which contain sets of integer values of cardinality 90, 900 and 9000, respectively. Instances were generated so the lengths of edges would get random values from a certain interval, independently for each criterion. All combinations of the intervals for the first (C1) and second (C2) criterion were checked. All the experiments from this group were performed on graphs with 20 vertexes. Both criteria were of *minisum* type.

For all problems, SPP, STP and TSP, three values were observed: upper bounds (UB), more precise upper bounds (PUB) and actual numbers of efficient points (EFF).

The upper bound was calculated by the formula

$$UB = \prod_{k=1}^{p} \left(n^U \theta_k - n^L \eta_k + 1 \right) \tag{6}$$

where n^L (n^U) is the lower (upper) bound of the number of variables that in feasible solutions may have value 1 and $[\theta_k, \eta_k]$ is the interval from which values for edges weights are taken (I1, I2 and I3 in the experiments). Although it is a rough bound, UB is polynomial respect to the problem size. It is easy to determine it only knowing the parameters of an instance.

The more precise upper bounds of the number of efficient points were calculated by the formula:

$$PUB = \min \left\{ f_1^2 - f_1^1, f_2^1 - f_2^2 \right\} + 1 \tag{7}$$

where $Y^k = \left(f_1^k, f_2^k \right)$, $k = 1,2$ are efficient marginal points obtained by lexicographic optimization where the k-th criterion has the first rank. PUB depends on the results of the experiments, and values Y^k used for its calculation are obtained from the average of 10 instances, as well as EFF.

The analysis and comparisons of the two kinds of upper bounds and the actual number of efficient points were performed in order to get an idea about order of magnitude and relations between the values of the UB, PUB and EFF. Results of this group of experiments are given in Table 1.

The second group of experiments was performed in order to check the dependency of the number of efficient points on the graph size. Because of the exponential complexity of the algorithms for finding all efficient points for all the problems, the experiments were performed for instances with up to 50 vertexes. In the earlier published experiments it was concluded that the number of efficient points for STP significantly more depends on the number

Table 1: Dependence of the upper bounds and efficient points number on the range of edge lengths for SPP, STP and TSP ($n = 20$)

prob:		SPP			STP			TSP		
C1	C2	UB	PUB	EFF	UB	PUB	EFF	UB	PUB	EFF
I1	I1	$4 \cdot 10^6$	136	9.6	$3 \cdot 10^6$	133	7.5	$3 \cdot 10^6$	483	40.5
I1	I2	$4 \cdot 10^7$	147	6.6	$3 \cdot 10^7$	172	8.9	$3 \cdot 10^7$	546	44.1
I1	I3	$4 \cdot 10^8$	166	7.6	$3 \cdot 10^8$	159	7.9	$3 \cdot 10^8$	498	39.3
I2	I2	$4 \cdot 10^8$	1027	6.5	$3 \cdot 10^8$	905	8.7	$3 \cdot 10^8$	4951	44.0
I2	I3	$4 \cdot 10^9$	1870	9.8	$3 \cdot 10^9$	1804	9.3	$3 \cdot 10^9$	5489	45.1
I3	I3	$4 \cdot 10^{10}$	13967	8.5	$3 \cdot 10^{10}$	13461	8.0	$3 \cdot 10^{10}$	47606	39.1
average				8.1			8.4			42.0

of terminal vertexes than on total number of vertexes in graph, so that problem was excluded from this group. Also, rough upper bound was not considered any more since in the previous experiments its value was many orders of magnitude bigger then the more precise upper bound. Here also the experiments were performed on instances with *minisum* criteria types.

Experiments results for this group are represented in Table 2.

Table 2: Dependence of the efficient points number on the graph size for SPP and TSP

problem	SPP		TSP	
v	PUB	EFF	PUB	EFF
10	622	3.5	771	4.7
20	1096	8.3	4809	43.9
30	2158	15.0	8365	105.5
40	2333	19.1	14116	223.5
50	2371	16.8	18521	373.9

The final group of experiments considered the types of criteria. Three combinations were observed:

- when both criteria were of type *minisum* (S/S),
- when the first criterion was of type *maximin* and second was of type *minisum* (M/S)and
- when both criteria were of type *maximin* (M/M).

This time the TSP problem was excluded from the experiments because the available software did not support solving TSP with *maximin* criterion. In addition, instances with both, 20 and 50 vertexes were observed. As in the previous group, only more precise upper bound and actual number of efficient points were considered.

The results of the third group of experiments are represented in Table 3.

Table 3: Dependence of the efficient points number on the type of criteria for SPP and STP

criteria type:		C=S/S		C=M/S		C=M/M	
problem	v	PUB	EFF	PUB	EFF	PUB	EFF
SPP	20	2886	8.1	938	5.4	309	2.6
	50	5689	20.0	1141	12.8	739	5.4
STP	20	2772	8.4	908	7.8	667	5.5
	50	2815	8.9	818	9.6	674	5.8

4 Conclusion

We can make the following conclusions which are based on results presented in Tables 1, 2 and 3.

Although the problems SPP, STP and TSP have very different nature, their number of efficient points show very similar characteristics.

The upper bounds given in columns UB, although have a polynomial growth with the problems size, represent a very rough bound, because of the big coefficients. Far more precise bounds are given in the columns PUB, but in order to obtain them, it is necessary to perform between p and p^2 optimizations per instance in order to obtain efficient marginal points (using lexicographic method).

The most interesting are the values in columns EFF. First of all, they are surprisingly small, and second, the influence of the observed parameters to it is very low or even insignificant.

Observing Table 1, it is obvious and expected that both upper bounds, UB and PUB, grow with the size of the intervals from which edges take their values. Very unexpected is that the actual number of efficient points does not show any dependence on the size of the intervals for all three problems.

It is also obvious that TSP has about five times bigger number of efficient points than SPP and STP which have similar number. Explanation is that for instances we used, TSP solutions contain more edges than solutions of SPP and STP. Consequences are that the distance between the two efficient marginal points is bigger, because of that PUB is also bigger and it is expected that bigger number of efficient points can be between them.

Observing the results from Table 2 a little deviation can be noticed for SPP between graphs with 40 and 50 vertexes. Namely, the number of efficient points on this stage starts to decrease. However, we concluded that the deviation is accidental, especially because in the next group of experiments, on different instances with the same characteristics (SPP, 50 vertexes, S/S) is obtained value 20.0 (shown in Table 3) which matches the value expected in Table 2. Here the number of efficient points is bigger for TSP and moreover, it grows faster than for STP. This is in accordance to the previous explanation because the number of edges in solution for TSP grows linearly with number of vertexes and for SPP is approximately $\sqrt{m}$.

Finally, the results from Table 3 show that the number of efficient points is smaller for *maximin* than for *minisum* type of criteria, i.e. if more criteria are of *maximin* type, smaller will be the number of efficient points. A slightly deviation of that rule is in the last row where for M/S combination of criteria types is a bigger number of efficient points than for S/S. Still we consider this as an accidental deviation.

It was mentioned before that the number of efficient points for STP does not depend much on the number of vertexes and it is also obvious from the last two rows of Table 3.

In all the considerations in this paper it was assumed that criteria are not correlated. On the other hand, the number of efficient points decreases with the increase of correlation between criteria. Since it is known that between many criteria used in practice correlation exists (the length, time and price of path, price and reliability etc.), we can expect even less number of efficient points when it comes to practical problems.

References

[1] M. Ehrgott, *Multicriteria optimization*, Springer-Verlag, 2000.

[2] M. Ehrgott, X. Gandibleux, "A survey and annotated bibliography of multiobjective combinatorial optimization", *OR Spektrum*, Vol. 22, pp. 425-46 2000.

[3] V.A. Emelichev, V.A. Perepelitsa, "On cardinality of the set of alternatives in discrete many-criterion problems", *Discrete Math. Appl.* Vol. 2, pp. 461-471, 1992.

[4] H.W. Hamacher, G. Ruhe, "On spanning tree problems with multiple objectives", *Ann. Oper. Res.*, Vol. 52, pp. 209-230, 1994.

[5] I.V. Sergienko, V.A. Perepelitsa, "Finding the set of alternatives in discrete multicriterion problems", *Cybernetics* Vol. 23, pp. 673-683, 1987.

[6] M. Visée, J. Teghem, M. Pirlot, E.L. Ulungu, "Two-phases method and branch and bound procedures to solve the bi-objective knapsack problem", *J. Glob. Optim.* , Vol. 12, pp. 139-155, 1998.

[7] M. Vujošević, M. Stanojević, "Multiobjective traveling salesman problem and a fuzzy set approach to solving it". In: D. Ivanchev, M.D. Todorov (eds), *Applications of Mathematics in Engeneering and Economics*, Heron Press, Sofia, pp. 111-118, 2002.

[8] M. Vujošević, M. Stanojević, "A bicriterion Steiner tree problem on graph", *Yugosl. J. Oper. Res.*, Vol. 13, pp. 25-33, 2003.

Milan Stanojević
University of Belgrade
Faculty of Organizational Sciences
154 Jove Ilića, 11000 Belgrade, Serbia
E-mail: milans@fon.bg.ac.yu

Mirko Vujošević
University of Belgrade
Faculty of Organizational Sciences
154 Jove Ilića, 11000 Belgrade, Serbia
E-mail: mirkov@fon.bg.ac.yu

Bogdana Stanojević
Transilvania University of Braşov
Department of Computer Science
50 Iuliu Maniu, Braşov, Romania
E-mail: bpop@unitbv.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 503-507

The Virtual Reconstruction of the Medieval Citadel of Suceava by Means of Virtual Reality Technologies

Beatrice Ştefănescu, Cătălin Nicolae Căruntu, Florin Iulian Jamt

Abstract: Virtual Reality is a relatively new IT domain. It proposes the immersion in a computer generated virtual world. A virtual reality application accesses the user's senses while offering a synthetic world, governed by preset rules.
There is presented an application that recreates a national monument from the North Eastern Romania (the Medieval Citadel of Suceava) by means of Virtual Reality Technologies. This application is a functional component of the Sim-Space - Multi-sensorial simulator for the navigation inside virtual environments, based on virtual reality technologies. Sim-Space is a software system used for implementing projects created by researchers from Suceava, Iasi and Bucharest. The Sim-Space system proposes assisting knowledge by experiment in the form of an active virtual trip in a medieval fortress from Moldova, the Suceava Medieval Citadel. It promotes Romanian culture and history education as a part of the European and universal heritage. Physical reconstruction of monuments in ruin is expensive and endangers the objectives' authenticity. Virtually restoring a cultural and historic monument based on old pictures and existing fragments solves cost problems while allowing almost limitless access.
Keywords: Virtual Reality, 3D worlds, virtual reconstruction

1 Introduction

Virtual Reality is defined as a new technology which stimulates the mind or senses to create a simulation of reality in the imagination. VR is a three-dimensional computer generated simulation in which the user is able to visualize and manipulate the contents of this environment [4], [5].

IBM defines virtual reality by purely technical criteria: "a human/computer interface which allows the user to experiment an interactive three-dimensional synthesis environment. This artificial world contains sounds and objects that simulate the real-world. The user, which can influence the virtual surrounding in real-time, dives in a synthetic environment that triggers the immersive experience" [3].

One of the precursors of VR research, the Romanian Grigore Burdea [3] defines the concept of virtual reality as a "3-I triad", a triangle having each side reported to the three fundaments of VR: Immersion, Interaction and Imagination.

Some stages of the practical realization of the Suceava Medieval Citadel virtual reconstruction application are presented, based on virtual reality technologies.

This application is a functional component of the Sim-Space - Multi-sensorial simulator for the navigation inside virtual environments, based on virtual reality technologies. Sim-Space is a software system used for implementing projects created by researchers from Suceava, Iasi and Bucharest. The Sim-Space system proposes assisting knowledge by experiment in the form of an active virtual trip in a medieval fortress from Moldova, the Suceava Medieval Citadel. It promotes Romanian culture and history education as a part of the European and universal heritage. Physical reconstruction of monuments in ruin is expensive and endangers the objectives' authenticity. Virtually restoring a cultural and historic monument based on old pictures and existing fragments solves cost problems while allowing almost limitless access [7].

The Sim-Space project proposes the realization of a multi-sensorial simulator to assist knowledge by experiment under the form of a active virtual trip in a medieval citadel from Moldova, The Suceava Crown Citadel. An argument in favor of this simulator development was the promotion of quality education and instruction in the domain of Romanian people's history and culture knowledge, as part of European and universal heritage. Physical reconstruction of monuments in ruin is expensive and endangers the objectives' authenticity. There is presented an application that recreates a national monument from the North Eastern Romania (the Medieval Citadel of Suceava) by means of Virtual Reality Technologies. This application is a functional component of the Sim-Space - Multi-sensorial simulator for the navigation inside virtual environments, based on virtual reality technologies. Sim-Space is a software system used for implementing projects created by researchers from Suceava, Iasi and Bucharest financed by private and public funds.

2 Modeling, development, assembling and rendering solutions

Computer systems aren't capable to manage and transfer the data traffic equivalent to the continual flux of stimuli currently received by the human senses. The reproduction of the real universe would completely saturate, as data traffic, the actual technological possibilities of the most powerful computers [3]. This is why the conception of a virtual world must be simplified, however, just to the point in which the virtual environment looses credibility and is no longer accepted by the subject. To avoid this, 3D graphic tricks are used.

The technical solutions used to model, develop, assemble and render the scenes and objects of the virtual reconstruction application are:

- modeling of objects and 3D scenes: Autodesk 3D Studio Max (good modeling and physical interaction between scenes and objects quality, compatibility with other assembling environments);
- plugin to achieve compatibility between object modeling and assembling environment and the 3D rendering engine: oFusion (exports objects created in 3D Sudio Max to Ogre3D format);
- 3D objects and scene assembling software: Blender and Irrlicht (open source solutions);
- 3D engine: Ogre 3D (open source scene oriented rendering engine).

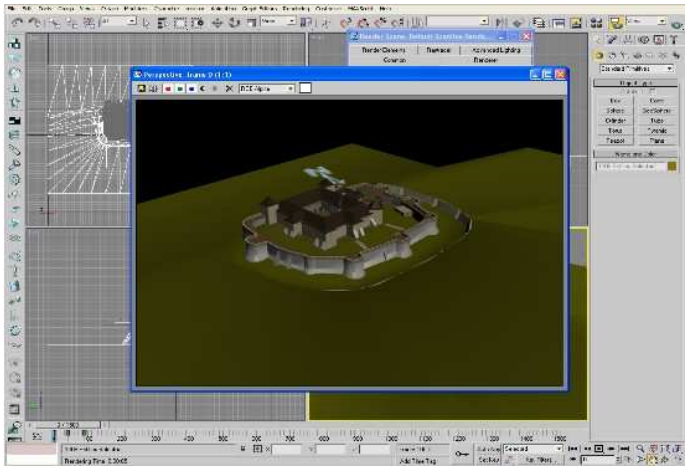


Figure 1: Complex scene 3D modeling and rendering. Medieval Citadel (draft)

As a demonstration, some 3D objects, models, scenes, textures are shown below, mentioning that all the over 100 objects and over 25 scenes developed in order to implement the Suceava Crown Citadel virtual reconstruction solution cannot be presented here. It is easy to display a simplified model with adequate performance. The situation of displaying millions polygons, even if they are loaded in memory, is not as simple [2], [6]. This balance will never be technically possible to achieve, because the faster the hardware, the more complex and refined the software models will be. Data structures were needed to be developed in order to permit the production of acceptable quality images, while keeping the resource consumption as low as possible.

3 Presentation of the used hardware equipment

Accessing virtual reality environments requires more resources than is required for an application running on a standard computer system. To assure the interaction of the application with the user, specific hardware devices were implemented: HMD (Head Mounted Display), motion tracker and VR Data Glove.

4 Virtual Reconstruction of the Suceava Crown Citadel

A scene is collection of objects, light sources and at least one camera position (to describe an angle of the environment visualization) [1]. Over 20 scenes were created in the form of geometrically defined object collections

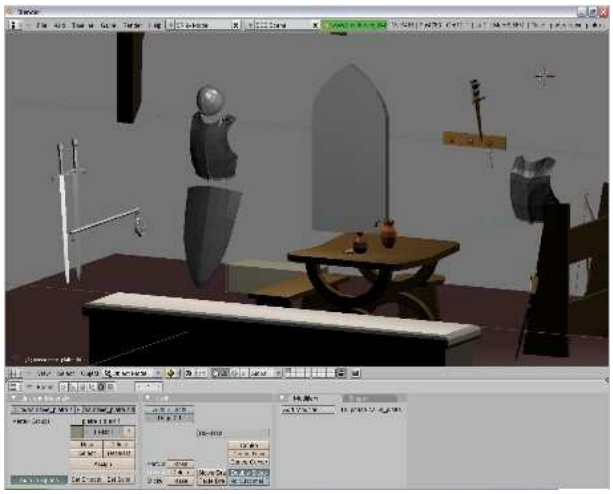


Figure 2: 3D scene and objects assembled in the Blender environment (example: Guard room)

Figure 3: Using i-Glasses 3D and VR DataGlove to test the application at SSIB Suceava

in the three-dimensional object-space together with the associated illumination and visualization parameters. The rendition is the process consisting in which an abstract description of a scene is converted into an image. By rendering, objects are illuminated and projected in the image space, where each of the pixels color intensity is calculated in order to obtain the visual effect of the final image.

3D image synthesis is a complex operation consisting in two stages: geometry processing and rastering (vectors to raster transformation). Entrance in the rendering pipeline is a description of the scene or database generated with the modeling. Objects are defined by geometrical parameters (scaled values or three-dimensional coordinates) and material properties (base colors and the way in which light is reflected ore refracted by the objects). Light sources are defined by position, intensity, color, and the cone of light.

5 Scene assembling and rendering

Objects generated after the modeling process using the 3D Studio Max environment are saved with *.MAX files format. In order to load these files in the Ogre3D created scenes, oFusion exporter is used. This process creates more mesh files and one material file for every exported mesh. Usage of suggestive names for these files is recommended to assist the objects management.

Although oFusion exports cameras with their associated lights, only meshes and materials will be exported, the cameras and lights being generated in the Ogre3D environment and in order to better manage the objects and scenes, the exporter is to copy the textures in the same directory as the other files. In order to avoid loading a large number of meshes, which would require supplementary code generation and rigorous object management,

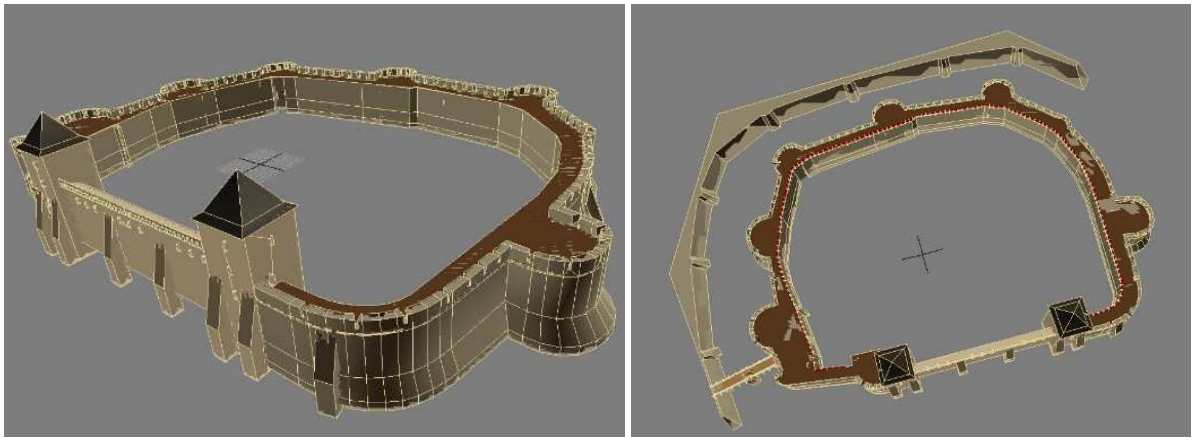


Figure 4: Simplified images depicting the Citadel realization stages

all immobile objects of the scene were grouped into a single larger object / scene (grouped objects cannot be individually moved, rotated or scaled; they form a single object to which all the operations would be applied). For example the initial scene of the Citadel's walls contained over 300 meshes. Beside the large code lines number and the difficult management of the loaded objects, the problem of their positioning in the scene must also be solved. Also, these objects make up only a part of the scene, many other meshes would be added (supplementary objects which would have to be loaded).

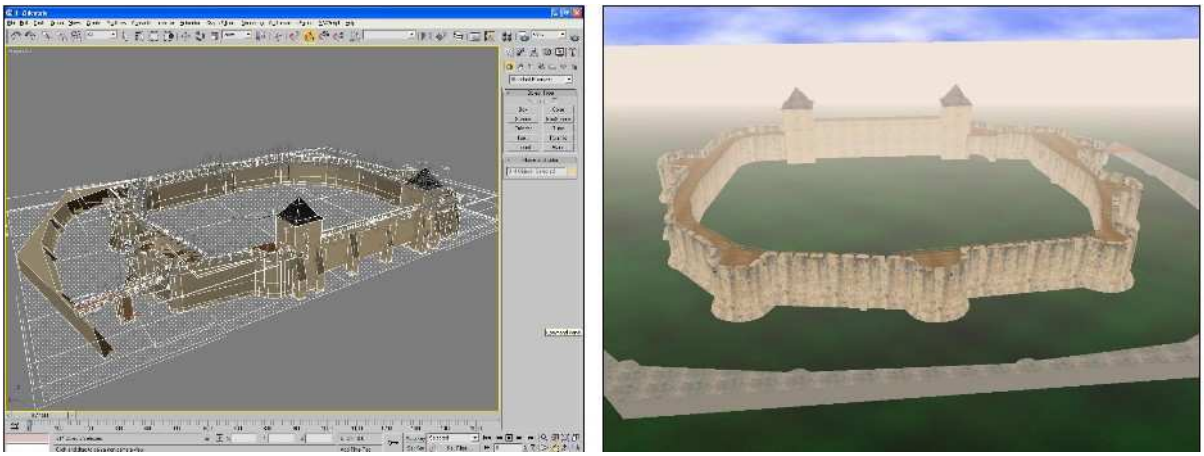


Figure 5: Citadel walls scene in 3D Studio Max (314 distinct object), then rendered by the Sim-Space simulator

In order to avoid positioning and scening orientations errors, all individually modeled objects are positioned in the same position, relative to the ground plane as they were placed in the modeling stages, all scaling, rotation and moving operations affecting the entire ensemble as a single object. After the mesh specification, the associated entity will be attached to the root node or any other node in the scene, which will thus become parent nodes, any action being reflected over their child nodes. The mesh file contains information referring to the material files, so the latter don't need specifications in the Ogre3D code.

An alternative solution using Macromedia Director was developed, in which collision detection is implemented.

6 Conclusions

The Sim-Space application is recommended as a Romanian research achievement, because it applies the modern concepts of virtual reality in the domain of culture and history knowledge about the Romanian people, as a part of universal and European heritage. The virtual reconstruction of the Medieval Citadel of Suceava is a functional application that offers data information economically accessible and by means of unconventional, persuasive

Figure 6: Soldiers resting quarters created in 3D Studio Max, then generated by the Sim-Space simulator

Figure 7: Chapel created in 3D Studio Max, then generated by the Sim-Space simulator

matters (to anybody, anywhere and anytime).

References

[1] Abraham F., Waldemar C., Cerqueira R., Campos J.L., "A Load Balancing Strategy for Sort-First Distributed Rendering", *Proceedings of SIBGRAPI* 2004.

[2] Akeley K., "Reality Engine Graphics. Computer Graphics Proceedings", *Conference Series, 1993*, ACM SIGGRAPH, July 1993, 109-116.

[3] Burdea G., Coiffet Ph., "La Realite Virtuelle", *Hachette*, Paris, 1995.

[4] Cadoz C., "Les Realites Virtuelles", *Dominos*, Flamarion, 1994, Paris.

[5] Chaillou C., "Architectures des Systemes pour la Synthese d'Images", *Dunod Informatique*.

[6] Crockett Th. W., "An Introduction to Parallel Rendering", *Parallel Computing*, Vol. 23, No. 7, July 1997, pp. 819-843.

[7] Filip F. Gh., Barbat B., "Informatica industriala. Noi paradigme si aplicatii", *Ed. Tehnica*, Bucuresti, 1999.

Beatrice Stefanescu, Catalin Nicolae Caruntu, Florin Iulian Jamt
SSIB SA
Romania, Suceava, Nicolae Balcescu Str. Nr. 1, CP:720066
E-mail: office@ssi-bucovina.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 508-511

Recognition Algorithm for Antenna-Free Graphs

Mihai Talmaciu, Elena Nechita

Abstract: During the last three decades, different types of decompositions have been processed in the field of graph theory. Among these we mention: decompositions based on the additivity of some characteristics of the graph, decompositions where the adjacency law between the subsets of the partition is known, decompositions where the subgraph induced by every subset of the partition must have predeterminate properties, as well as combinations of such decompositions.
In various problems in graph theory, for example in the construction of recognition algorithms, frequently appears the so-called weakly decomposition of graphs.
In this paper we introduce the notion of quasi-weakly decomposition of a graph G, we show the existence of a quasi-weakly decomposition, depending on the existence of the weakly decomposition. In addition, we present a recognition algorithm for the class of co-antenna-free graphs.
Keywords: Co-antenna-free graph, weakly decomposition, quasi-weakly decomposition, recognition algorithm.

1 Introduction

Throughout this paper, $G = (V,E)$ is a connected, finite and undirected graph, without loops and multiple edges, having $V = V(G)$ as the vertex set and $E = E(G)$ as the set of edges. $\overline{G}$ (or $c-G$) is the complement of G. If $U \subseteq V$, by $G(U)$ we denote the subgraph of G induced by U. By $G-X$ we mean the subgraph $G(V-X)$, whenever $X \subseteq V$, but we simply write $G-v$, when $X = \{v\}$. If $e = xy$ is an edge of a graph G, then x and y are adjacent, while x and e are incident, as are y and e. If $xy \in E$, we also use $x \sim y$, and $x \not\sim y$ whenever x,y are not adjacent in G. A vertex $z \in V$ distinguishes the non-adjacent vertices $x,y \in V$ if $zx \in E$ and $zy \notin E$. If $A,B \subset V$ are disjoint and $ab \in E$ for every $a \in A$ and $b \in B$, we say that A,B are *totally adjacent* and we denote by $A \sim B$, while by $A \not\sim B$ we mean that no edge of G joins some vertex of A to a vertex from B and, in this case, we say that A and B are *non-adjacent*.

The *neighbourhood* of the vertex $v \in V$ is the set $N_G(v) = \{u \in V : uv \in E\}$, while $N_G[v] = N_G(v) \cup \{v\}$; we simply write $N(v)$ and $N[v]$, when G appears clearly from the context. The neighbourhood of the vertex v in the complement of G will be denoted by $\overline{N}(v)$.

If $N[v] = V$, then v is called a *dominating vertex* in G. If $D \subset V$ and every vertex from $V-D$ has at least one neighbour in D, then D is called a *dominating set* of G. If $D \subset V$ and $\overline{N}_G(D) \neq \emptyset$, then D is a *non-dominating set* of G.

The neighbourhood of $S \subset V$ is the set $N(S) = \cup_{v \in S} N(v) - S$ and $N[S] = S \cup N(S)$. A *clique* is a subset Q of V with the property that $G(Q)$ is complete. The *clique number* of G, denoted by $\omega(G)$, is the size of the maximum clique.

By P_n, C_n, K_n we mean a chordless path on $n \geq 3$ vertices, a chordless cycle on $n \geq 3$ vertices, and a complete graph on $n \geq 1$ vertices, respectively.

A graph is called $triangulated$ if it does not contain chordless cycles having the length greater or equal to four.

A *antenna* graph is isomorphic to $G = (\{a,b,c,d.e.f\}, \{af, fd, fe, db, ec, bc\})$.

Let F denote a family of graphs. A graph G is called *F-free* if none of its subgraphs is in F. The *Zykov sum* of the graphs G_1, G_2 is the graph $G = G_1 + G_2$ having:

$$V(G) = V(G_1) \cup V(G_2),$$
$$E(G) = E(G_1) \cup E(G_2) \cup \{uv : u \in V(G_1), v \in V(G_2)\}.$$

When searching for recognition algorithms, frequently appears a type of partition for the set of vertices in three classes A,B,C, which we call a *weakly decomposition*, such that: A induces a connected subgraph, C is totally adjacent to B, while C and A are totally nonadjacent.

The structure of the paper is the following. In Section 2 we recall the notion of weakly decomposition, and we define the notion a quasi-weakly decomposition. In Section 3 we establish the existence of a quasi-weakly decomposition in a graph G. In Section 4 we give a recognition algorithm for the class of co-antenna-free graphs.

2 Preliminary results

At first, we recall the notions of weakly component and weakly decomposition, and then we define the notion of quasi-weakly decomposition.

Definition 1. ([2], [5], [6]) *A set $A \subset V(G)$ is called a weakly set of the graph G if $N_G(A) \neq V(G) - A$ and $G(A)$ is connected. If A is a weakly set, maximal with respect to set inclusion, then $G(A)$ is called a weakly component. For simplicity, the weakly component $G(A)$ will be denoted with A.*

Definition 2. ([2], [5], [6]) *Let $G = (V,E)$ be a connected and non-complete graph. If A is a weakly set, then the partition $\{A, N(A), V - A \cup N(A)\}$ is called a weakly decomposition of G with respect to A.*

Definition 3. *Let $G = (V,E)$ be a connected and non-complete graph. A partition $(\{v\}, X - \{v\}, Y, Z)$, where v is a non-dominating vertex, $Z = \overline{N}(v)$, $X \cup Y = N[v]$, $v \in X$, $Z \sim Y$, $Z \not\sim X$, is called a quasi-weakly decomposition of the graph G.*

Remark. Since G is connected and v is a non-dominating vertex, it follows that, in Definition 3, always $Z \neq \emptyset$ and $Y \neq \emptyset$.

If $X = \{v\}$ is a quasi-weakly decomposition, then we have, in fact, a complete bipartition, namely $\{\{v\} \cup \overline{N}(v), N(v)\}$. The question that naturally arises is: when a quasi-weakly decomposition of G does exist? The answer is given in Proposition 1.

The name of "*quasi-weakly decomposition*" is explained in what follows. A set $D \subset V$ is a dominating set of G, if every vertex in $V - D$ has at least one neighbor in D. A set $D \subset V$ is a non-dominating set of G if $\overline{N}_G(D) \neq \emptyset$. The sets X and Z are non-dominating, because $X \not\sim Z$, but Y is non-dominating if there is at least a vertex of X that is non-adjacent with a vertex of Y (X, Y, Z are the elements of the quasi-weakly decomposition). This would justify the name of almost non-dominating decomposition. In fact, a quasi-weakly decomposition $(\{v\}, X - \{v\}, Y, Z)$ is the weakly decomposition (X, Y, Z) in which v is a dominating vertex in the subgraph induced by $X \cup Y$. Moreover, in the quasi-weakly decomposition, the subgraph induced by X is connected. This is the reason why we adopt, in what follows, the name of "*quasi-weakly decomposition*".

Below we remind a characterization of the weakly decomposition of a graph.

The name of "*weakly component*" is justified by the following result.

Theorem 1. ([3], [5], [6]) *Every connected and non-complete graph $G = (V,E)$ admits a weakly component A such that $G(V - A) = G(N(A)) + G(\overline{N}(A))$.*

Theorem 2. ([5], [6]) *Let $G = (V,E)$ be a connected and non-complete graph and $A \subset V$. Then A is a weakly component of G if and only if $G(A)$ is connected and $N(A) \sim \overline{N}(A)$.*

The next result, that follows from Theorem 1, ensures the existence of a weakly decomposition in a connected and non-complete graph.

Corollary 1. *If $G = (V,E)$ is a connected and non-complete graph, then V admits a weakly decomposition (A, B, C), such that $G(A)$ is a weakly component and $G(V - A) = G(B) + G(C)$.*

Theorem 2 provides an $O(n + m)$ algorithm for building a weakly decomposition for a non-complete and connected graph.

Algorithm for the weakly decomposition of a graph ([5])
Input: A connected graph with at least two nonadjacent vertices, $G = (V,E)$.
Output: A partition $V = (A, N, R)$ such that $G(A)$ is connected, $N = N(A)$, $A \not\sim R = \overline{N}(A)$.

begin
  A := any set of vertices such that
  $A \cup N(A) \neq V$
  $N := N(A)$
  $R := V - A \cup N(A)$
  while ($\exists n \in N$, $\exists r \in R$ such that $nr \notin E$) *do*
    begin
      $A := A \cup \{n\}$
      $N := (N - \{n\}) \cup (N(n) \cap R)$
      $R := R - (N(n) \cap R)$
    end
end

3 The quasi-weakly decomposition of a graph

In this section we present some sufficient conditions for the existence of a quasi-weakly decomposition and an algorithm for finding such a decomposition.

The next result, that follows from Theorem 1, ensures the existence of a quasi-weakly decomposition.

Corollary 2. *If $G=(V,E)$ is a connected, non-complete graph then V admits a quasi-weakly decomposition $(\{v\},A-\{v\},B,C)$, such that (A,B,C)* is a weakly decomposition, *$G(A)$ is a weakly component and $G(V-A)=G(B)+G(C)$, and v is a dominating vertex in $G(V-C)$.*

The existence of a dominating vertex in $G-C$, where (A,B,C) is a weakly decomposition, is ensured by the following result.

Propozitia 1. *Let $G=(V,E)$ be a connected non-complete, P_4-free graph and (A,N,R) a weakly decomposition of G with $G(A)$ as a weakly component. If $G(A)$ is a C_4-free graph then G admits a quasi-weakly decomposition.*

4 Some applications of the quasi-weakly decomposition

In this section, using the quasi-weakly decomposition, we present a recognition algorithm for the co-antenna-free graphs. Firstly, we give a characterization of a co-antenna-free graph.

Theorem 4. *Let $G=(V,E)$ be a connected graph, v a dominating vertex in $G-Z$ and $(\{v\},X-\{v\},Y,Z)$ is a quasi-weakly decomposition with $N(X-\{v\})=\{v\}\cup Y$. Then G is co-antanna-free if and only if the following assertions are true:*
(i) $G-X$, $G-Z$ are co-antenna-free graph;
(ii) there exists no $2K_2$ in $G((X-\{v\})\cup Y)$ with endpoints of the edges in both sets of the quasi-weakly decomposition (which means that each edge has an endpoint in $X-\{v\}$ and other one in Y) and there also is not $2K_2$ in $G(Y)$;
(iii) there exists no y of Y, midpoint of only one P_4 with both endpoints either in $X-\{v\}$ or in Y, so that $N_G(y)\cap(X-\{v\})$ is a clique.

Proof. Let G be a connected graph, v a dominating vertex in $G-Z$ and $(\{v\},X-\{v\},Y,Z)$ is a quasi-weakly decomposition with $N(X-\{v\})=\{v\}\cup Y$. If G is a co-antenna-free graph then $G-X$ and $G-Z$ are co-antenna free graph. If there exists $2K_2$ in $G((X-\{v\})\cup Y)$ more specific x_1y_1, $x_2y_2\in E$, where $x_1,x_2\in X-\{v\}$ and $y_1,y_2\in Y$ then $G(\{v,x_1,x_2,y_1,y_2,z\})$ is co-antenna, $\forall z\in Z$. Because $N(X-\{v\})=\{v\}\cup Y$, result for any vertex of Y there is an adjacent vertex from $X-\{v\}$. If there are two nonadjacent vertices $x_1,x_2\in Y$, of a $2K_2$ ($x_1y_1,x_2y_2\in E$) from $G(Y)$ which share a neighbor, x, from $X-\{v\}$ then $G(\{z,x_1,x_2,y_1,y_2,x\})$, $\forall z\in Z$, is a co-antenna because $Z\sim Y$ and $Z\not\sim X-\{v\}$. We suppose that any two vertices that are non-adjacent x_1,x_2 of any $2K_2$ from $G(Y)$ have neighbors in $X-\{v\}$ two non-adjacent vertices: $ax_1,bx_2\in E$, $a,b\in X-\{v\}$. Since $\{v\}\sim X-\{v\}$ and $Z\not\sim X-\{v\}$, $Z\sim Y$, results $G(\{v,a,b,x_1,x_2,z\})$ is a co-antenna, $\forall z\in Z$. Results *ii*). We suppose that *iii*) doesn't hold. Then *a*) or *b*) hold, where:
a) there is $a,d,c\in X-\{v\}$; $b\in Y$ and $cd\notin E$, $ad\notin E$, $ad\in E$;
b) there is $b,d\in X-\{v\}$, $a,c\in Y$ and $bd\in E$.
In both cases, $G(\{a,b,c,d,y,z\})$ is co-antenna, $\forall z\in Z$.
We suppose that i), ii) and iii) holds and that, however, there is a induced subgraph $G(A)$ of G, co-antenna, where $A=\{x,x_1,x_2,y_1,y_2,z\}$, $xx_1,xx_2,xy_1,xy_2,x_1y_1,x_2y_2,y_1z,y_2z\in E$. $A\cap X\neq\emptyset$, else $G-X$ is not co-antenna free. $A\cap Z\neq\emptyset$, else $G-Z$ is not co-antenna free. $A\cap Y\neq\emptyset$, else does not $X\not\sim Z$. From $G-X$, $G-Z$ co-antenna free follows that $G(X\cup Z)$, $G(Y\cup Z)$ are co-antenna free. If $x\in X-\{v\}$ then if $x_1,x_2\in X-\{v\}$ and $y_1,y_2\in Y$ then ii) is contradicted, and in the other cases either $A\subseteq X\cup Y$ or a adjacent relationship between the elements of A is incorrect ($zy_1\notin E$, $zx_1\in E$). If $x\in Y$ then $z\notin Z$. If $z\in X-\{v\}$ then $y_1y_2\in(X-\{v\})\cup Y$. If either $y_1,y_2\in X-\{v\}$ or $y_1,y_2\in Y$ then $A\subseteq X\cup Y$ otherwise (which means that $y_1\in X-\{v\}$ and $y_2\in Y$) ii) is contradicted. If $z\in Y$ then $x_1,x_2\in(X-\{v\})\cup Y$. If either $x_1,x_2\in X-\{v\}$ or $x_1,x_2\in Y$ then $A\subseteq X\cup Y$ else (which means that $x_1\in X-\{v\}$ and $x_2\in Y$) iii) is contradicted. If $x\in Z$ then $z\in X\cup Z$, $x_1,x_2\in Y\cup Z$, $y_1,y_2\in Y\cup Z$. If $z\in Z$ then $A\subseteq Y\cup Z$ otherwise either $G(\{z,x_1,x_2,y_1,y_2,x\})$ is a co-antenna or an adjacent relationship between the elements of A is incorrect ($x_2y_1\in E$).

The above theorem leads to the following recognition algorithm.

Input: $G=(V,E)$ a connected graph satisfying the conditions in Theorem 4

Output: An answer to the question: "Is G a $co-antenna$-free graph" ?

begin

1. $L = G$; // L is a list of graphs
2. *while* $(L \neq \emptyset)$
   begin
a. extract an element H from L;
b. find a quasi-weakly decomposition $(\{v\}, X-\{v\}, Y, Z)$ for H;
c. *if* ii) or iii) don't holds the G is co-antenna
   else introduce in L the connected, non-complete components of
   $G-X$, $G-Z$
   Return: "G is $co-antenna-free$"
   end
end

In what follows, we give some remarks on the algorithm.
Step 2b) is $O(n+m)$.
Step 2c) is $O(nm)$:
A recognition algorithm for $2K_2$-free graphs is polynomial, according [7]
The test from iii) from the above theorem, refering to the fact if $T = N(y) \cap (X - \{v\})$ is clique, is this:
$K := \{y\}$
for each $v \in T$ do
if $\{v\} \sim K$ then
$K := K \cup \{v\}$
This is (nm).
The test iii) from the above theorem hold this way:
If P_4 does not exist // this is linear times, see [1] and [6] *else*
If the neighborhood of a middle of P_4 in $X - \{v\}$ is clique // this is $O(nm)$.
Globally, steps 1 and 2 execute $O(n^2m)$.

References

[1] D.G. Corneil, Y. Perl and L.K. Stewart Burlinham, *A Linear Recognition Algorithm for Cographs*, SIAM J. Computing 14 (4), 1985, pp. 926-934.

[2] C. Croitoru, M. Talmaciu, *A new graph search algorithm and some applications*, presented at ROSYCS 2000, Univ. "Al.I.Cuza" Iaşi, 2000.

[3] C. Croitoru, E. Olaru, M. Talmaciu, *Confidentially connected graphs*, The annals of the University "Dunarea de Jos" of Galati, Suppliment to Tome XVIII (XXII) 2000, Proceedings of the international conference "The risk in contemporany economy".

[4] R.Lin and S.Olariu, *An Nc Recognition Algorithm for Cograph*, Journal of Parallel and Distributed Computing, 13, 76-91 (1991).

[5] M. Talmaciu, *Decomposition Problems in the Graph Theory with Applications in Combinatorial Optimization - Ph. D. Thesis*, University "Al. I. Cuza" Iasi, Romania, 2002.

[6] M. Talmaciu, E. Nechita, *Recognition Algorithm for diamond-free graphs*, INFORMATICA, 2007, Vol. 18, No. 3, 457-462, ISSN 0868-4952.

[7] ***, *antenna-free graphs*, http://wwwteo.informatik.uni-rostok.de/isgci/classes/gc-394.html.

Mihai Talmaciu and Elena Nechita
University of Bacău, Department of Mathematics and Informatics
Bacău, St.Spiru Haret no.8
E-mail: mihaitalmaciu@yahoo.com, elenechita@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 512-515

Dynamic Distribution Model in Distributed Database

Leon Ţâmbulea, Manuela Horvat

Abstract: A database can be stored in a centralized manner, or its fragments can be distributed in a set of nodes in a network [1]. Some fragments can be stored in a single node, or in more than one node (are replicated)[2]. The operations required by client (as reading, adding, updating, or deleting records) could be performed over the fragments according to the client rights for the mentioned fragment and the fragment rights (read rights and write rights) in the node where the operation is required.
The distribution of the fragments in the database nodes and the rights allocation can be performed in a static mode by specifying exactly the rights and placement for each fragment, or can be performed in a dynamic mode taking into account factors as: the node rules for fragment management or the requests flow in the distributed database.
In the current paper is proposed a method for a dynamic distribution of the fragments in the nodes of a distributed database.
Keywords: : distributed database, dynamic model

1 Introduction

Let's note with C the distributed data collection (a distributed database) stored in the nodes of a network. The C collection can be considered as a set of data fragments, and such a fragment can be found (replicated) in more than one node. In the model it's considered that the set of fragments (the collection C is composed of) is fixed, and their distribution in the network's nodes can be changed dynamically.[3]

Let's note with

$$C(N) \subseteq C$$

the set of fragments contained by collection C that are located on the network node N.

More than one user have access to the data fragment D located on the node N, and the access rights can be different [4]. If at least one user has the write permission for the fragment D stored in the node N, than the fragment D has the write permission in the node N. If all the users have only read permission s on fragment D located in the node N, than the fragment D has the read permission in the node N. The management of a fragment D having write permission is more expensive than the management of the same fragment having read permission (because a writing operation in the fragment D must be propagated to all the fragment's replicas).

A data fragment D and a node N can be in a one of the following relations on a specific moment of time:

- the fragment D is not stored in the node N
- the fragment D is stored in the node N and has the read right for this node, so it can only be read from the node N.
- the fragment D is stored in the node N and has the write right for this node, so it can be read and updated in the node N.

This relation between a fragment D and a node N can be dynamically changed according to a set of factors. This model of dynamic allocation proposes a method for allocating dynamically the access permissions for a fragment D and a node N.

2 Requests Resolving

2.1 Read requests

A client read request q for a set of data in the database (an sql select request for example) is sent to a node N. This request require data from several fragments:

$$r(q) = \{D_i, i \in I\}$$

If all the fragments are stored on the node N

$$r(q) \subseteq C(N)$$

then the response for the request q can be determined in the node N.

If some fragments in the r(q) relation are not stored on the node N, then an algorithm is used to determine the response for request. The algorithm is specific to the distributed database system management, and some algorithms can be mentioned [2]:

- The fragments from r(q) are transferred in the node N, and then the answer is determined.
- The request q is passed to a node N1, where is executed, and the answer is sent back from node N1 to the originating node (node N for this case).
- The request q is split into sub queries/requests, these sub-queries are executed in parallel in different nodes, and the results are sent to the originating node (node N) where the answer to the initial request q is determined.

2.2 Write requests

If a client sends an updating or adding request q to a node N, then the write operations in different fragments

$$r(q) = \{D_i, i \in I\}$$

will be performed before the initial request is finalized. The protocol for distributed transactions execution is used for completion of an updating operation for a fragment.

If all the fragments referred by the r(q) relation are stored in the node N and have the write permission in the node N, then the update request is completed in this node.

If there are fragments referred by the r(q) relation that are not stored in the node N and do not have the write permission, than:

- the request can be sent to other node
- if certain conditions are valid, the fragments with read only permissions stored in the node N can change permissions: from the read right to write right
- for the fragments that are not stored in the node N, the update request is sent to the node N1 that stores the primary copy/replica of the fragment, and the request becomes distributed. The node N recives a copy of fragment D asyncronously (after the write transaction was completed), the copy on the node N will have the read permission.

3 Dynamic reallocation of the fragments

A collection C (a distributed database) is assumed to use and determine some statistical information stored in the database "catalog" such as:

- **WMax** the maximum number of nodes where a fragment can be stored with write permission. The greatest number that can be set to this parameter is the maximum number of nodes. An optimal value must be found for this parameter, as a too large number of write permission would have impact on the time for updating requests to complete.
- **WMix** the minimum number of nodes where a fragment must be stored with write permission. This parameter must be at least 1 (needed for a writing request to be executed), but for backup reasons it is recommended to be greater or equal with two.
- **WM(N)** the minimum number of write request for a fragment D in a node N, (it makes no importance if the fragment is stored or not on the node N), is used when analyzing if the fragment should get the write permission on the node N.

- A fragment D having a specific permission should be stored on a node N only if the node N received requests for accessing (for reading or for writing) the fragment D. The state of a fragment D stored on the node N is analyzed after monitoring it for period of time **P(N)**. The period of time **P(N)** can be constant for all the nodes in the collection C or can be specific for each node.
- In each node can be defined a function/condition *cond(N, D, Pa)* that would decide if a fragment D can be stored in the node N with the access permission Pa (read, write). This function/condition is set determined according to a set of factors such as: the available memory on the node N, the current loading, etc.
- For each fragment D are stored the nodes where the fragments copies are located and the fragments permissions in these nodes.
- Statistical information computed in the node N for a specific period of time (24 hours, one hour, five minutes). The statistical information is updated each time a data set/fragment is accessed for reading or writing.
  - **R(N, D)** is the number of read requests received by the node N, for accessing the fragment D stored on the node.
  - **W(N, D)** is the number of write requests received by the node N, for accessing the fragment D stored on the node.
  - **W1(N, D)** is the number of the write requests received by the node N, for accessing the fragment D which is not stored on the node.
  - **W(D)** is the number of the nodes where the fragment D is stored having write permission.

A request q can be split into elementary queries (operations) that execute over different fragments stored in the node N. If such an elementary request (a join operator for example) uses fragments that are not stored in the same node, than all the fragments should be transferred to the same node.

The node N receives a request q and in order to execute it, it needs the access to a fragment D. For resolving the access and transfer problem, the following algorithm can be defined:

- If the fragment
$$D \in C(N)$$
and the right of the fragment D stored in the node N corresponds to the operation required by the request, than the request q can be executed in the node N, and the values R(N, D) and W(N, D) are incremented by one (according to operation type: read or write).
- If the fragment
$$notD \in C(N)$$
then:
  - If the request requires a read operation, than based on the database catalog are determined the nodes where a replica of the fragment can be found and a node N1 is chosen. The required fragment is transferred from the node N1 to the node N for the request q to be completed. If the condition *cond(N, D, read)* is true, then the fragment D will be stored on the node N having read right and the value of **R(N, D)** will be initialized with 1: **R(N, D) = 1**. Asynchronously, the database catalog should be updated with the information regarding the other nodes in the database that store the fragment D.
  - In the write operation is required by the request, and *cond(N, D, write)* is true, then a replica of the fragment D is transferred on the node N, and the request q is completed.
  - The request q is sent from the node N to the node N1 that stores the fragment D with write permission, so q can be completed.

Using the events that are raised in the node N, such as the expiration of the period of time set in the P(N) parameter, or if some conditions are valid (the available space is below a specified minimum limit), a background process for database clearing can be executed. The process is based on a simple algorithm:

- If a fragment D stored in a node N, having read right has the number of read accesses equal to zero **(R(N, D) = 0)**, then the fragment D is removed from the node N, and the database catalog is updated by removing the information related to fragment D.

- If a fragment D stored in the node N, having write right has the number of write accesses equal to zero **(W(N, D) = 0**), and the **W(D)** $>$ **WMin** is true (there are enough copies of the fragment D with write permission in the database), then the fragment D is removed from the node N, and the database catalog is updated by removing the information related to fragment D.
- If **(W(D)** $<$ **WMax)** and **(cond(N, D, write) is true)** then the fragment D will be copied in the node N from a replica with write permission will get the write permission. The database catalog is updated with the information related to fragment D.

Observation: the third step in this algorithm can be placed also at the step 2c in the precedent algorithm.

4 Conclusion

This article proposes a simple method of dynamic distribution of the fragments of a distributed database and a method for allocation the access rights for these fragments. The method implementation is simple due to the fact that the data required by the management process (algorithm) are stored in the database catalog.

The future work is to create and to execute different simulations tests in order to determine the optimal values for the algorithms parameters.

References

[1] M. Tamer Õzsu, *Distributed & Parallel DBMS* , 2003.

[2] Ţâmbulea Leon, *Baze de Date* , 2003.

[3] Maurice Herlihy, Victor Luchangco, Mark Moir, and William N. Scherer, "III. Software Transactional Memory for Dynamic-Sized Data Structures", *Proceedings of the Twenty-Second Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC)*, pp. 92-101, 2003.

[4] Tim Harris, Simon Marlow, Simon Peyton Jones, and Maurice Herlihy, "Composable Memory Transactions", *ACM Conference on Principles and Practice of Parallel Programming 2005* , 2005.

Manuela Horvat, Leon Ţâmbulea
Babeş Bolyai University, Cluj-Napoca
Faculty of Mathematics and Computer Science
Department of Computer Science
400084, M. Kogalniceanu St., 1, Cluj-Napoca, Romania
E-mail: manuela.petrescu@gmail.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 516-520

Practical Security Issues for a Real Case Application

Florin Vancea, Codruţa Vancea

Abstract: Security techniques for application software and systems are well-known and often used in various situations. They range from encryption to special data handling and good usage practices. However, providing proper levels of security for a complete application is not simple and requires careful design and implementation. The paper presents some design decisions and implementation aspects for a real case application used for distributing confidential payroll information to a large number of employees via browser-based technology.
Keywords: Security, best practices, example

1 Introduction

Modern application software is facing multiple challenges. Many of them stem from the wide introduction of networking technology. As network communication became readily available the original model of stand-alone application has shifted to the client-server architecture. The original network-based thick client model is now replaced by the thin client model with multi-layered applications, due to the low management and maintenance overhead of this model.

The new model offers instant easy access to application functionality to any workstation that has a web browser and network access. This wide availability increases significantly the potential number of users and requires careful design of the application components in order to maintain a good level of performance. From a security point of view and compared to the single workstation model, this thin client multi-layer model introduces multiple segments of responsibility, each carrying its own specific security problems and requirements.

We will discuss some of the issues outlined above in relation to a real case application, namely a system that should distribute in secure manner personal payroll information to a large number of employees using existing general-purpose workstations connected to enterprise intranet. The application was designed, coded and implemented for a real beneficiary. For obvious reasons names, locations, number figures and some details will be omitted.

2 Application requirements and environment

A large company with offices distributed all over the country has a significant number of employees. The purpose of the application (further called Application) is to distribute on demand personal payroll information to all the employees in a fast and cost-effective way while maintaining a high level of confidentiality and trust. The application should be as easy to use as possible, without sacrificing security, because the computer "literacy" level of the employees may be low.

The cost-effective way chosen was to allow each employee to access a web-based application using existing computers with a common browser. Access to the server is possible through company intranet. This is good, because it limits the security threats compared to Internet-based access. However, the network is not considered safe because the company intranet is wide and complex enough to allow good chances for an eavesdropper.

For obvious reasons, the access should be identified and authenticated by a pre-known username and a personal password. The pre-known attribute is particularly important because at the moment the application was launched in production all users had to be notified in a secure manner about the username and the initial password. For verification purposes, the username has to be pre-known and chosen to be some uniform and comfortable identifier of the employee (the employee Human Resource identifier printed on the last paper pay slip).

All payroll data is generated within the Payroll system which has a very limited number of users and a strongly restricted access policy. For the purposes of this paper we will consider Payroll as a strong link and will not discuss it further. On a monthly basis the whole set of data computed for all employees is exported from Payroll and imported into the Application. As soon as the import is over, data should be available to all employees. Each employee should be able to access the Application and obtain its own and only its own subset of payroll information, for the current month and for up to 6 previous months. The general Application operating environment is represented in Figure 1.

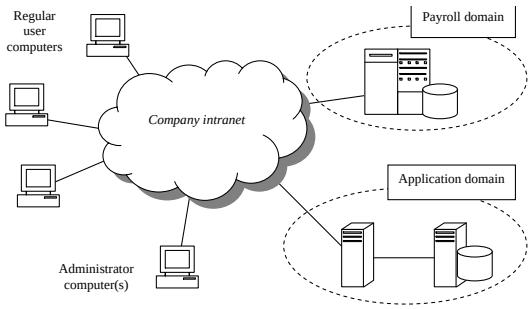


Figure 1: Application environment

These requirements are not uncommon at all and a plain careful implementation should cover all of them without any problem. However, the Application is a bit more special in that it aims to protect the confidentiality of the payroll information even against persons holding special privileges, like the database administrator, the application administrator, the network maintenance staff or the application developer. Of course, there are some persons that we cannot fully protect the data against, either because it makes no sense (e.g. the Payroll limited set of users who already have access to the information) or because it is practically impossible (e.g. the operating system administrator for the server). In the case of the latter we did try to make the access hard enough, but a software-only system has no real defense against a debugger.

3 Application architecture and common security measures

The Application has a multi-tier structure illustrated in figure 2.

The Application data is stored in an Oracle database backend. For the purpose of this analysis it can be a dedicated server or the Application data can be simply hosted in one schema of a multi-purpose server. The protection mechanisms we provide should ensure proper security even in the latter case. The operating system of the database server is of no relevance to our analysis, as we will see shortly. The middle layer is based on J2EE technology and is composed of application modules running inside a JBoss container. The server machine that runs it is using a Linux operating system. On client side, Internet Explorer is considered to be the common browser, but there is nothing that would stop the user to access the system with any other browser as long as it has the required SSL capabilities.

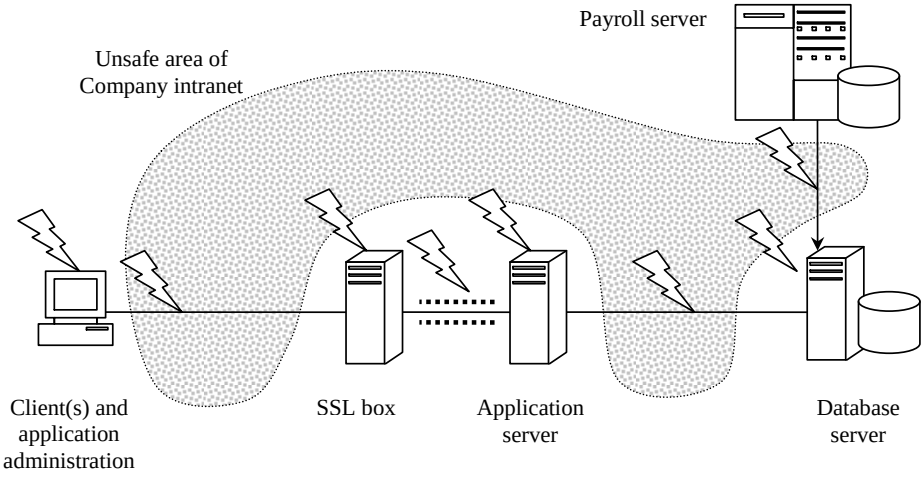


Figure 2: Elements of the Application and potential threats

Figure 2 shows also the communication paths between the system components and the various places having specific security risks. Mostly all communication is done using the company intranet which is considered to be insecure against eavesdropping. For reasons explained below, SSL protection and the application server are decoupled in separate entities, communicating over a private connection. From left to right, the security threats identified are:

1. client-level threats
2. eavesdropping and man-in-the-middle attacks over the client-application link
3. attacks directed to SSL server
4. eavesdropping and man-in-the-middle attacks on the private link
5. attacks directed to the application server
6. eavesdropping on the link between the application server and application database server
7. data theft and manipulation directly in database server
8. eavesdropping on the link between Payroll and application database server

At client level there is not much we can do to protect the data without sacrificing functionality. Some measures were taken however, like instructing the browser not to cache sensitive pages. The rest had to be left mostly at user's discretion. Obviously, users with sensitive roles are specially trained to avoid bad security practices.

The client access link is protected using SSL. To avoid man-in-the-middle attacks, server's certificate was deployed to all workstations in the company using existing proven and secured computer administration techniques. From the security point of view providing a SSL access point directly at the application server level would be the best choice. Unfortunately, from a performance point of view this solution was not acceptable. The application server can provide SSL access but being Java-based this would have placed a significant and unacceptable load on the application server itself. The problem is aggravated by the fact that the Application is used with a very uneven pattern. During each entire month there is little or no usage except for the pay day when all users try to access the system over a very short period of time. To maintain good performance levels without over sizing the application server (thus keeping the cost at a minimum) the SSL protection was decoupled to a dedicated box. Originally the plan was to use a dedicated hardware SSL box for maximum performance and security but later a dedicated software solution was used due to lower cost.

The link between the SSL box and the application server is carrying unencrypted traffic, including user data to be protected and more important passwords and passphrases used to authenticate the user and to unlock special features accessible to special users. This is a sensitive resource and should be accessible only to someone already trusted. In the final implementation, the SSL box and the application server are each running in a separate virtual machine and both are located in the same physical machine managed by one person - the infrastructure administrator. This way the private link described above is not a physical one but a virtual one and is not exposed to direct wiretap risks.

The application server is exposed to multiple threats and discussing them all is not possible within the limited scope of this paper.

The link between the application server and the database server is exposed. One method to protect it would have been to use Oracle's secure link features. However, since we will protect the sensitive data inside the database, too, all sensitive data carried by this link is going to be encrypted at application level. Protecting this link is optional, but may improve overall security. The same notes apply to the link between the Payroll system and the application database.

The database itself is exposed, too. The assumption we made is that data inside the database is exposed to multiple risks. To cover for all those risks, data should be transmitted and stored in encrypted form and good key management is required. There are some native security features that Oracle offers, but those would have significantly increased the cost.

4 Particular security measures

Primary data is generated inside the Payroll system. This is considered to be a secure black box. Monthly data set is generated there and then encrypted using a key generated for each export. Encrypted data is copied

then to a separate schema on application database server (different than the Application schema) and the generated key is delivered to a Payroll trusted operator who has also the importer role in the Application. Upon operator's request, the Application will decrypt the Payroll data, re-encrypt it properly for long-term storage then delete the temporary encrypted Payroll data. This minimizes the time window available to attack the Payroll data during the export-import procedure.

Re-encryption of sensitive data is carried by the application server. Given the large amount of data, using a single key is not acceptable, so several operational keys are used. Each user's data for a particular month is protected with symmetric encryption using a single key. This ensures good performance levels by limiting the encryption setup overhead for a particular user's data. The key to be used is randomly chosen from a set of eligible "fresh" keys.

All encryption keys have to be stored somewhere safely. Storing them directly in the database is obviously unsafe and we wanted to avoid storing them in a file on the application server, too, because we had to protect them against the application administrator and the application developer. Therefore, those keys are stored in the database but encrypted with a symmetrical cipher. When the application is initialized all operational keys are retrieved and decrypted and they remain in plaintext only inside application server memory. The master key protecting other keys is required only for a short time until the operational keys are retrieved and is not stored in memory in any way.

This scheme presents some potential problems. First, the master key protecting the operational keys should be protected, too. Storing it hidden inside the application itself is not a good choice, so we decided to let a special user manage the security of this key. Of course, giving this special user the master key directly is a risk because this user may gain access to the database and use the master key to retrieve ultimately all the encrypted data.

Second, a unique master key may be lost and then all data becomes inaccessible. Therefore we store several sets of encrypted versions of the operational keys, each set protected by a different master key belonging to a different special user. Protection of each master key is achieved by storing it in the database, too, but associated with the user identity and protected by encryption using as key a combination between a user-known passphrase and a server-known key. Since there is a clear separation of domain access and indirect holders of master keys do not have normally access to the application server or the database the true key is secure.

Having several indirect holders of master keys is good, but we had to provide a safe mechanism for introducing a new trusted indirect holder with its master key. Basically, introducing a new holder means generating a new master key, storing in the database a new set of encrypted operational keys protected with this master key and finally storing the new master key protected by the passphrase of the new holder. This operation can only be triggered by en existing master key holder that has also a secure role of "introducer".

Key lifetime must be controlled. After one operational key is used for a while the key should be "retired". That key will not be used for new encryptions but it has to be kept around for recovering old data. New keys may be generated, but we cannot store them in the database without the master key, or actually without enough master keys to guarantee the system's reliability. Therefore, generating new keys will require at least two key holders to collude and enter their passphrase within a limited time window. Newly generated keys will be propagated to other key holders whenever the keys are in memory and some key holder enters his/hers passphrase.

A more subtle issue is that only proper keys should be placed in application memory. If fake keys can be injected in the Application internal key pool then normal retrieve operation is not compromised. However, if an import is performed in such conditions new user data may be encrypted with the fake key(s) and get exposed. When the system has already some properly encrypted keys in the database, adding fake keys is impossible. The initialization procedure makes sure that all keys present in the set are properly decrypted with the same master key. This does not prevent the possibility of building an entire set of fake keys, all protected with a fake master key owned by a fake identity and injecting them in the unsafe database, then re-initializing the Application with the fake set. The retrieve functionality will be lost (no proper old keys in the fake set) but any import performed in this situation would compromise new data.

The solution is to create "trusted roles" and "trusted accounts" using a chain of trust protected by public-key cryptography. The Application identifies and authenticates users by a classical method (stored hashes of passwords and names) but this is highly vulnerable to data manipulation in the database. For sensitive operations (like key generation or data import) additional identification and authentication is required.

The authenticity of the database records holding the trusted user and role information is guaranteed by digital signature. The Application has a root trusted user that owns some initial trusted roles with a grant option. All other trusted users have their database records signed by the root trusted user or by an intermediate introducer. To protect against manipulation the records are both upwards and downwards linked and the links are protected by the

signature.

Using this scheme, everything is only as secure as the root trusted user records. Some attacker may delete the entire tree of trust together with all the keys and replace it with its own tree and its own keys. The manipulation is exactly equivalent to normal initial database initialization and cannot be avoided. To protect against this kind of attack, the root trusted user records are signed with a private key held by the application developer. The Application itself contains already the developer's public key because it's required for checking the integrity of the code, so it can check the integrity of the root trusted user records. With this mechanism in place, the initialization follows this path:

1. The new root trusted user chooses a passphrase. The system generates a public-private key pair and stores it in the database protected by a key derived from the passphrase.
2. The public key is sent to the application developer who builds the record for a new trusted role and signs it with its own private key, then sends the result to the root trusted user.
3. The root trusted user enters the received signed message in the system, thus gaining a new trusted role, as specified by the developer.
4. The process is repeated for all required trusted roles.
5. Once the process is over, the root trusted user can introduce new trusted users and grant them new trusted roles, all properly double-chained and signed with its private key.
6. For any trusted user, once the user enters its passphrase the system validates that the recovered private/public keys are a correct pair and that the trust double-linked chain is validated up to the known developer public key
7. Two of the trusted users holding the trusted role "master key holder" generate the initial set of keys.
8. The system is ready for the first import, performed by a trusted user holding the trusted role "importer".

The common user has to be protected against hash manipulation, too. This goal is achieved by encrypting the user password and a salt value with one of the operational keys. At login time, the value is decrypted and the password is checked. The salt value is the same salt used when encrypting this user's monthly secure data. This tweak is not offering any additional true security but protects against another kind of manipulation of references to secure data.

5 Summary and Conclusions

Designing a properly secured application is more than applying encryption to data and we believe and our experience confirms that this is not yet fully understood by many application developers and system administrators. When designing a good solution one should take into account many other variables like the existing environment, available resources and technology, business requirements and cost constraints. Beyond good security practices, one should be careful about the whole picture when considering threats and finding the best solutions may require considering some tradeoffs.

References

[1] B. Schneier, *Applied Cryptography,* John Wiley & Sons, 1996.

[2] Sun Microsystems, *Java JCE "Java Cryptography Extension",* Available at: http://www.javasoft.com/products/jce/

Florin Vancea, Codruţa Vancea
University of Oradea
Department of Computers, Department E.M.U.E.E.
5 Armatei Romane Street, Oradea, Romania
E-mail: fvancea@uoradea.ro, cvancea@uoradea.ro

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 521-525

Balanced PID Tuning Application to Series Cascade Control Systems

Ramon Vilanova, Orlando Arrieta

Abstract: This communication provides an approach for the application of PID controllers within a cascade control system configuration. Based on considerations about the expected operating modes of both controllers, the tuning of both inner and outer loop controllers are selected accordingly. This fact motivates the use of a tuning that for the secondary controller that provides a balanced set-point / load-disturbance performance. A new approach is also provided for the assimilation of the inner closed-loop transfer function to a suitable form for tuning of the outer controller. Due to the fact that this inevitably introduces unmodelled dynamics into the design of the primary controller, a robust tuning is needed.
Keywords: PID Control, Cascade control systems

1 Introduction

The introduction and use of an additional sensor that allows for a separation of the fast and slow dynamics of the process results in a nested loop configuration as it is shown in figure (1). Each loop has associated its corresponding PID controller. The controller of the inner loop is called the secondary controller whereas the controller of the outer loop as the primary controller, being the output of the primary loop the variable of interest. The rationale behind this configuration is that the fast dynamics of the inner loop will provide faster disturbance attenuation and minimize the possible effect disturbance before they affect the primary output. This set up involves two controllers. It is therefore needed to tune both PIDs. The usual approach involves the tuning of the secondary controller while setting the primary controller in manual mode. On a second step, the primary controller is tuned by considering the secondary controller acting on the inner loop. It is therefore a more complicated design procedure than that of a standard single-loop based PID control system.

In this paper a design issue that has not been addressed is considered: the *tradeoff* between the performance for set-point and load-disturbance response. When a load -disturbance occurs at the primary loop, the global load-disturbance depends on the set-point tracking performance of the secondary loop. In addition, good load-disturbance performance is expected for the secondary controller in order to attenuate disturbances that enter directly at the secondary loop. Also, it is well known that when the controller is optimally tuned for set-point response, the load-disturbance performance can be very poor [1]. Based on this observation this paper proposes the use of a balanced performance tuning [2] for the secondary loop. Furthermore, an approximation procedure is provided in order to assimilate the dynamics seen by the primary controller to a First-Order-Plus-Time-Delay model such that usual tuning rules for PID control can be applied. However here a robust tuning is suggested, because, the primary controller will need to face with unmodelled dynamics coming from the model approximation used for the secondary loop. Note that this kind of approximation is always needed if simple-model based tuning rules are to be applied.

The rest of the paper is organized as follows. Next section presents the cascade control configuration and control setup to be used. Section 3 provides the main contribution of the paper as the design approach involving tuning of the controllers and approximation method. Section 4 presents an application example whereas section 5 ends with some conclusions and suggestions for further research.

2 Cascade Control

A typical configuration for cascade control is shown in figure (1), where an inner loop is originated from the introduction of an additional sensor in order to separate, as much as possible, the process fast and slow dynamics. As a result the control system configuration has an inner controller $C_2(s)$ with inner loop process $G_2(s)$ and outer loop controller $C_1(s)$ with outer loop process $G_1(s)$. Disturbance can enter at two possible distinct points: d_1 and d_2. The rationale behind this configuration is to be able to compensate for the best, possible disturbance d_2, before it is reflected to the outer loop output. In order to accomplish that purpose it is essential that the inner loop exhibits a faster dynamics that allows for such early compensation.

According to this, the overall process $G(s) = G_1(s)G_2(s)$ is split into the two parts $G_1(s)$ and $G_2(s)$ and associated controllers. The two controllers are standard feedback controllers that are assumed to take the usual ISA-PID form as:

This work has received financial support from the Spanish CICYT program under grant DPI2007-63356. Also, the financial support from the University of Costa Rica and from the MICIT and CONICIT of the Government of the Republic of Costa Rica is greatly appreciated.

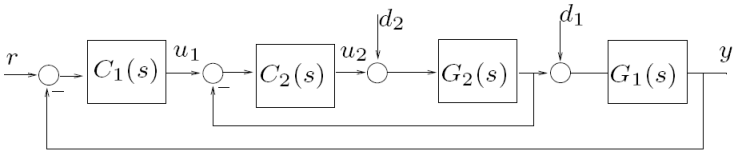


Figure 1: Cascade Control Configuration

$$C_j(s) = K_{pj}\left(1 + \frac{1}{T_{ij}s} + \frac{T_{dj}s}{1+(T_{dj}/N_j)s}\right) \tag{1}$$

where K_{p1} and K_{p2} are the proportional gains, T_{i1} and T_{i2} the integral times, T_{d1} and T_{d2} the derivative times and finally N_1 and N_2 the derivative time noise filter constants.

3 Approach for Cascade Control Design

The proposed approach for cascade control design is presented according to the different design stages needed in order to completely determine all the control system components.

3.1 Inner loop and outer loop process models

A description of the inner process is assumed to be available as a First-Order-Plus-Time-Delay (FOPTD) model. Well known procedures [3] can be applied in order to provide such approximation. Therefore $G_2(s)$ is assumed to obey to $G_2(s) = \frac{K_2}{1+T_2 s}e^{-L_2 s}$. Along similar lines, considering the output of $G_2(s)$ as the input to $G_1(s)$ a model of the same characteristics can be obtained for the slow dynamics part of the system as $G_1(s) = \frac{K_1}{1+T_1 s}e^{-L_1 s}$. The FOPTD process model is widely used and constitutes the starting point for many of the existing PID tuning procedures.

3.2 Inner loop controller tuning

According to the role of $C_2(s)$ within the control system configuration it is important to bear in mind that a good disturbance rejection is expected for the inner loop (in order to accommodate the possible disturbance entering at d_2) as well as good set-point tracking capabilities. Effectively, when a disturbance enters at d_1 or when a reference change occurs, the outer loop controller $C_1(s)$ will generate the corresponding reference change for the inner loop controller $C_2(s)$. On the other side it is well known [1] that if we tune $C_2(s)$ for good disturbance rejection (set-point response) the set-point (load-disturbance) response can degrade considerably. Therefore, a balanced tuning is needed for the inner loop controller where a *tradeoff* for both operation modes is considered. These kind of issues have been introduced in [2] and provided a tuning approach for suboptimal PID design called γ-tuning. This γ-tuning is based on the definition of a performance degradation index that takes into account how the response degrades with respect to the optimal one. Due to paper length constraint the method is not reproduced here. For a detailed description see [1] and [2].

3.3 Model for Outer loop tuning

Once the tuning of the secondary loop has been completed, the effective system, $G_e(s)$, seen by the outer loop can be determined.

$$G_e(s) = H_2(s)G_1(s) = \frac{C_2(s)G_2(s)}{1+C_2(s)G_2(s)}G_1(s) \tag{2}$$

where the complete expression for $H_2(s)$ takes the form

$$H_2(s) = \frac{p_2(s)e^{-L_2 s}}{p_1(s)+p_2(s)e^{-L_2 s}} \tag{3}$$

with,

$$p_1(s) = (T_2 s + 1) T_{i2} s \left(\frac{T_{d2}}{N_2} s + 1\right) \tag{4}$$

$$p_2(s) = K_2 K_{p2} \left(1 + (T_{i2} + \frac{T_{d2}}{N_2}) s + \frac{T_{i2} T_{d2}}{N_2} (N_2 + 1) s^2\right) \tag{5}$$

where the presence in the denominator of the irrational term $e^{-L_2 s}$ difficultly the obtention of the effective process model $G_e(s)$ and posterior outer loop controller design. In this paper it is proposes to perform an approximation of the denominator of $H_2(s)$ on the basis of a new polynomial $p(s)$ and delay term $e^{-\theta s}$ such that $p_1(s) + p_2(s) e^{-L_2 s} \approx d(s) = p(s) e^{-\theta s}$. First of all the required order of $p(s)$ is determined. Note the effective transfer function $G_e(s)$ can be rewritten as

$$G_e(s) = \frac{p_2(s) e^{-L_2 s}}{p(s) e^{-\theta s}} \frac{K_1 e^{-L_1 s}}{1 + T_1 s} = \frac{p_2(s)}{p(s)} \frac{K_1 e^{(-L_1 - L_2 + \theta) s}}{1 + T_1 s} \tag{6}$$

The purpose here is to have an approximation to $G_e(s)$ on the basis of a FOPTD. Therefore, a second order polynomial $p(s)$ is needed. Note this way $G_e(s)$ will behave with a -20dB roll-off as a first order system. On the other hand the approximation with a first order system will make no sense. On this basis, the first step is to determine the coefficients of $p(s) = p_o + p_1 s + p_2 s^2$ and the value of θ. Secondly, to find:

$$\tilde{G}_e(s) = \frac{K_e e^{-L_e s}}{T_e s + 1} \approx G_e(s) \tag{7}$$

as the FOPTD model the design of the outer loop controller will be based on. Note that $L_e = L_1 + L_2 - \theta$ and $K_e = K_1$ (integral action in the inner loop will assure $H_2(0) = 1$). Therefore it only rest to determine the vale of T_e.

The approximation by using $d(s)$ is performed according to the following procedure. First of all the first three terms of the McLaurin expansions of both $p(s) e^{-\theta s}$ and $p_1(s) + p_2(s) e^{-L_2 s}$ are performed. Equating corresponding terms provides the following set of equations:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ -\theta & 1 & 0 \\ \theta^2 & -2\theta & 2 \end{pmatrix}}_{A_\theta} \underbrace{\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix}}_{\overrightarrow{p}} = \underbrace{\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}}_{\overrightarrow{b}} \tag{8}$$

with,

$$\begin{pmatrix} b_0 \\ b_1 \\ \\ b_2 \end{pmatrix} = \begin{pmatrix} K_2 K_{p2} \\ T_{i2} + K_2 K_{p2} (T_{i2} + T_{d2}/N_2 - L_2) \\ 2 T_{i2} T_{d2}/N_2 + 2 T_{i2} T_2 + \\ K_2 K_{p2} (L_2^2 - 2 L_2 (T_{i2} + T_{d2}/N_2) + \\ 2 T_{i2} T_{d2} (N_2 + 1)/N_2) \end{pmatrix} \tag{9}$$

Once a value for θ is provided, $\overrightarrow{p}$ can be determined by $\overrightarrow{p} = A_\theta^{-1} \overrightarrow{b}$. On the other hand, the value of θ will be determined in such a way that provides the better approximation; $d(s)$; in the following sense:

$$\min_{\theta, \overrightarrow{p} = A_\theta^{-1} \overrightarrow{b}} \left\| p_1(s) + p_2(s) e^{-L_2 s} - p(s) e^{-\theta s} \right\|_\infty \tag{10}$$

Once the values for θ and $\overrightarrow{p}$ are got, the determination of T_e in (7) is right straightforward.

3.4 Outer loop controller tuning

Once the model for the effective process (7) is available we can proceed with the tuning of the outer loop controller. As the inner loop will provide compensation for local disturbances, we can think of the outer loop controller to be tuned in order to accommodate good performance for the set-point response. Bearing in mind that the process model used for design comes from an approximation of an higher order dynamics, aggressive tunings should be avoided. A very simple and FOPTD model based tuning rule that guarantees some degree of robustness is provided in [4]. On the basis of the effective model approximation, this tuning rule reads:

$$\begin{aligned} K_{p1} &= \frac{T_{i1}}{2.65K_eL_e} \\ T_{i1} &= T_e + 0.03L_e \\ \frac{T_{d1}}{N_1} &= 1.72L_e \\ N_1 + 1 &= \frac{T_e}{T_{i1}} \end{aligned} \tag{11}$$

4 Example

The presented approach is now exemplified by means of a simulation example. Consider the following definitions for the process models:

$$G_1(s) = \frac{1}{100s+1}e^{-40s} \qquad G_2(s) = \frac{5}{20s+1}e^{-4s} \tag{12}$$

The tuning of the secondary controller has been performed by application of the optimal ISE tuning rules [5] for set-point tracking operation $K_{p2}^{sp} = 0.89$, $T_{i2}^{sp} = 17.83$, $T_{d2}^{sp} = 2.34$ and load-disturbance $K_{p2}^{ld} = 1.40$, $T_{i2}^{ld} = 5.34$, $T_{d2}^{ld} = 2.39$. Application of the γ-tuning [2] with a weighting factor that gives a 25% extra weight to the load-disturbance performance degradation with respect to the set-point performance degradation ($W_{ld} = 1.25$ and $W_{sp} = 1$) provides $K_{p2}^{\gamma} = 1.15$, $T_{i2}^{\gamma} = 11.59$, $T_{d2}^{\gamma} = 2.37$. For all these tunings the value of the derivative time filter is taken as $N_2 = 10$. Consequently, for each one of the three secondary loops, an effective model approximation, $\tilde{G}_e(s)$ is computed: $K_e^{sp} = 1$, $T_e^{sp} = 99.78$, $L_e^{sp} = 44$; $K_e^{ld} = 1$, $T_e^{ld} = 96.58$, $L_e^{ld} = 44$; $K_e^{\gamma} = 1$, $T_e^{\gamma} = 97.73$, $L_e^{\gamma} = 44$; and corresponding outer loop tunings generated as: $K_{p1}^{sp} = 0.87$, $T_{i1}^{sp} = 101.10$, $T_{d1}^{sp}/N_1^{sp} = 99$; $K_{p1}^{ld} = 0.84$, $T_{i1}^{ld} = 97.90$, $T_{d1}^{ld}/N_1^{ld} = 102$; $K_{p1}^{\gamma} = 0.85$, $T_{i1}^{\gamma} = 99.05$, $T_{d1}^{\gamma}/N_1^{\gamma} = 101$. Figure (2) shows the performance with respect to a step reference change is almost identical for the three scenarios. However, the disturbance attenuation to a load disturbance entering at d_1 is clearly superior for the load disturbance optimally tuned secondary controller. Note for the γ-tuning the load-disturbance is also clearly better than that of the set-point tuning (without loosing performance with respect to a step change).

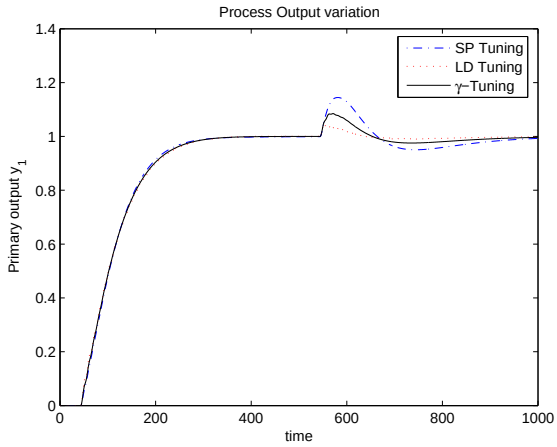


Figure 2: Primary output for a step in r and d_2

However, the noted lower robustness margins of the load-disturbance tuning generate a system that may be very sensitive to model errors [2]. If we assume, for example, a 5% uncertainty in the secondary process time-constant the performance for the set-point and γ-tuning is maintained whereas for the load-disturbance the system is critically unstable as it is shown in figure (3).

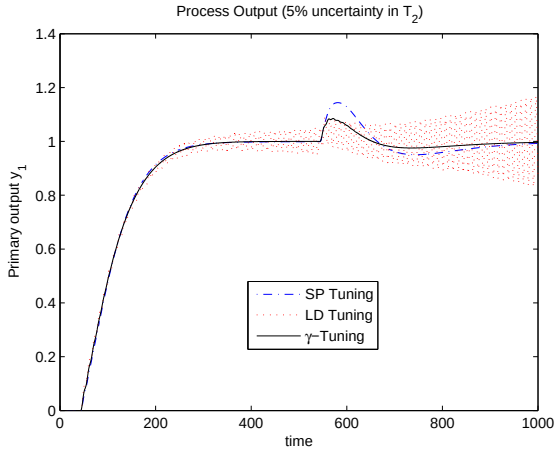


Figure 3: Primary output assuming a 5% uncertainty in T_2

5 Conclusions

This paper has addressed the problem of PID controller tuning within a cascade control system configuration. A procedure has been outlined that considers a balanced operation (set-point and load-disturbance) for the secondary controller and a robust tuning for the primary controller. In order to facilitate the design of the primary controller an approximation method has been provided that generates a FOPTD approximation suitable for PID tuning. The success of the approx has been shown by means of an example. However a more deep analysis has to be done, specially with respect to the obtention of concrete uncertainty bounds for the secondary loop modelling.

References

[1] O. Arrieta and R. Vilanova, "Performance degradation analysis of Optimal PID settings and Servo/Regulation *tradeoff* tuning," *CSC07, Conference on Systems and Control, Marrakech-Morocco*, 2007.

[2] O. Arrieta and R. Vilanova, "Servo/regulation *tradeoff* tuning of PID controllers with a robustness consideration," *CDC07, 46th IEEE Conference on Decision and Control, New Orleans, Louisiana-USA*, 2007.

[3] M.A. Johnson and M. H. Moradi, *PID Control. New Identificaton and Design Methods*, Springer Verlag, 2005.

[4] R. Vilanova, "IMC based robust PID design: Tuning guidelines and automatic tuning," *Journal of Process Control*, vol. 18, pp. 61–70, 2008.

[5] M. Zhuang and D. Atherton, "Automatic tuning of optimum PID controllers," vol. 140, no. 3, pp. 216–224, 1993.

Ramon Vilanova, Orlando Arrieta
Universitat Autònoma de Barcelona
Department of Telecommunication and Systems Engineering
ETSE, Campus UAB, 08193, Bellaterra, Barcelona
E-mail: {Ramon.Vilanova, Orlando.Arrieta}@uab.cat

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 526-530

Expected Interaction Based Design Oriented Frequency Domain Stability Condition for Decentralized Control of TITO System

Ramon Vilanova

Abstract: This paper analyzes process and control interaction in multivariable two-input two-output processes. It is well known that the off-diagonal terms in the matrix transfer functions introduce interaction effects that are to be taken into account at the controller design stage. When decentralized control is to be used, it is of primary importance to have knowledge of the constraints imposed by local designs into the behavior of the overall system. Special attention has to be paid to the potential instability caused by interaction effects generated by the control action on the other loop. This paper provides an analysis by means of an interaction measure defined in terms of the joint effect of process and control interaction.

Keywords: Multivariable Systems, Interaction, Decentralized control

1 Introduction

The use of full multivariable control approaches has reached an advanced stage of maturity (Model Predictive Control is a clear example [1]). However, the use of a decentralized control structure still remains the more widely used structure in the process industries. The main reasons of handling a Multiple Input Multiple Output (MIMO) system by means of multiloop Single-Input Single-Output (SISO) controllers are the simple controller structure that results and the fact that loop failure is easily handled. However the presence of interactions among the loops introduce an inherent difficulty to the design of these local controllers. In the presence of strong interactions the effectiveness of the decentralized controllers can be seriously deteriorated or even cause instability.

Among MIMO systems, the most common form encountered in process industries is the Two-Input Two-Output (TITO) case and the purpose of this paper is to introduce a simple interaction measure that takes into account the desired dynamics for the local closed-loops therefore aimed at measuring the final interaction that will be present on the resulting decentralized control system. As interaction finally depends on the other-loop selected controller, some information on the expected behavior of the other loop will be needed. This will allow us to establish a link between the desired dynamics for the independent closed-loops; generated interaction; and potential closed-loop instability.

The rest of the paper is organized as follows. Next section reviews the interaction problem in decentralized control of a TITO system. Section 3 presents the proposed interaction measure and a related index for expected closed loop performance and condition for stability. Section 4 presents and example of application and in section 5 conclusions are drawn as well as some points where actual research is conducted.

2 Decentraliced control and closed-loop Interaction

Let a Two-Input Two-Output (TITO) process be represented by the following matrix transfer function

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = = G(s) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{1}$$

The off-diagonal terms $G_{12}(s)$ and $G_{21}(s)$ cause coupling between the two separate input-output pairs. This interaction effect between loops is recognized to generate several undesirable effects for control design and it is necessary to take it into account and incorporate some measure of such interaction at design stage.

Assume a diagonal pairing 1-1/2-2 is applied (on the other case the process transfer matrix can be conveniently rearranged) and a controller $K_1(s)$ is applied to control y_1 by using u_1, and a controller $K_2(s)$ to control y_2 by using u_2 as shown in figure (1). If we close the loops around $G_{11}(s)$ and $G_{22}(s)$ respectively, it would be natural to state the design of each controller in terms of these open-loop diagonal terms and try to find $K_1(s)$ and $K_2(s)$ such that

$$T_i(s) = \frac{K_i(s)G_{ii}(s)}{1 + K_i(s)G_{ii}(s)} \approx T_i^d(s) \tag{2}$$

This work has received financial support from the Spanish CICYT program under grant DPI2007-63356

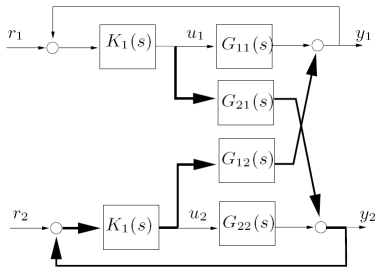


Figure 1: Decentralized control of a TITO process

where $T_i^d(s)$ specify the desired closed-loop dynamics for the corresponding closed-loops. We will refer to these loops as the independent local-loops. However, under such decentralized control strategy, the off-diagonal interaction terms will introduce additional dynamics that will cause the transfer functions *seen* by the controllers to be, respectively:

$$G_i(s) = G_{ii}(s) - \frac{K_j(s)G_{ij}(s)G_{ji}(s)}{1+K_j(s)G_{jj}(s)} \tag{3}$$

These relations clearly show the tuning of one of the controllers depends on the tuning of the other one. This kind of interaction is the main problem that arises in decentralized control. A better look at the interaction terms can be obtained by introducing $T_j(s)$ from (2) into (3) obtaining:

$$G_i(s) \quad = \quad G_{ii}(s)\left(1-\frac{G_{ij}(s)G_{ji}(s)}{G_{ii}(s)G_{jj}(s)}T_j(s)\right) = G_{ii}(s)\left(1-I(s)T_j(s)\right) \tag{4}$$

where

$$I(s) = \frac{G_{12}(s)G_{21}(s)}{G_{11}(s)G_{22}(s)} \tag{5}$$

is the Rijnsdrop interaction measure [7]. Among all these, otherwise classical, interaction measures, it is the $I(s)$ index that enters the transfer function seen by each local controller in a more direct way. Therefore preferred. Expression (4) reveals the total interaction results from a combination of the inherent (open-loop) process interaction as measured by the Rijnsdrop index and the control interaction expressed by the independent closed-loop transfer functions. It seems therefore not complete to take any decision on the pairing of manipulated and control variables as well as the design of the local controllers without taking control interaction into account.

3 Interaction Measure and Interaction effects Analysis

From relation (4) it is clear that a *deadlock* is present: the local second loop has to be known in order to know the control interaction in the first loop and the other way round. The only way of performing an independent loop analysis is to introduce some information about the designed controller for the other loop. This will result in a sequential loop closing approach (see [8] and [9] for example) that has some recognized disadvantages [10] such as the need for an iterative design as the final controller may depend on the order the loops are being closed. Instead, we propose here to use the desired closed loop dynamics in order to know how much interaction is to be expected from the other loop. This is equivalent to assume we are able to design $K_i(s)$ such that $T_i(s) = T_i^d(s)$. The following interaction terms are respectively associated to loop1 and loop2 are respectively:

$$I_1(s) = I(s)T_2^d(s) \qquad I_2(s) = I(s)T_1^d(s) \tag{6}$$

This way, each independent loop will have associated its interaction measure. The interaction terms defined in (6) are interpreted as the *expected* total interaction that would be generated if the desired dynamics for the local control loops are achieved. According to (4) these interaction levels provide a modification of the original process transfer functions that can be interpreted in terms of a multiplicative uncertainty formulation $G_i(s) =$

$G_{ii}(s)(1-I_i(s))$. Therefore the well known constraint for Robust Stability [11] apply[1]. Taking $G_{ii}(s)$ as the nominal systems for design, and assuming the local controllers $K_i(s)$ are designed such that $T_i^d(s)$ are achieved, both stability constraints are expressed as $\|T_i^d(s)I_i(s)\|_\infty < 1$. That turn out to be identical for both loops and imply a first important constraint on the closed loop dynamics that can be specified for each independent loop as:

$$\|T_1^d(s)T_2^d(s)\|_\infty < 1/\|I(s)\|_\infty \tag{7}$$

Note that even the desired closed loop dynamics will generally not be completely achieved, constraint (7) may be taken as a reference for potential instability in terms of the desired dynamics and should be taken into account prior to the design of the local controllers. If (7) is not satisfied it must be taken as an indication of a not well-posed desired dynamics specification. Note this constraint may be somewhat conservative as it arises from only considering a magnitude bound on the model perturbing term. In addition, within a Robust Control framework, the model perturbation term uses to be a magnitude bounded uncertain term. Here, as we know $T_i^d(s)$ as well as $I(s)$, the constraint can be stated more precisely as

$$|T_1^d(jw)T_2^d(jw)I(jw)|_{w=w_{pc}} < 1 \tag{8}$$

where w_{pc} is the phase crossover frequency. However, the real closed loop relations; $T_i^r(s)$ between $y_i - u_i$ will result from $K_i(s)$ closing the loop around $G_i(s)$. These transfer functions are related to the desired ones as follows:

$$T_i^r(s) = \frac{K_i(s)G_i(s)}{1+K_i(s)G_i(s)} = \frac{K_i(s)G_{ii}(s)(1-I_i(s))}{1+K_i(s)G_{ii}(s)(1-I_i(s))} = T_i^d(s)\frac{1-I_i(s)}{1-T_i^d(s)I_i(s))} \tag{9}$$

From where it can de stated that it will be possible to get $T_i^r(s) \approx T_i^d(s)$ if

1. $I_i(s) \approx 0$. There is no interaction coming from the other loop.
2. $T_i^d(s) \approx 1$. Therefore at the frequency range where perfect control is specified. This will enter in conflict with the stability constraint (7) and (8) and no perfect control would be specified nor achieved over all the frequency range.

4 Example

Let us consider the following example taken from [12]:

$$G(s) = \begin{pmatrix} \frac{1}{s+0.1} & \frac{1}{s+0.1} \\ \frac{2}{s+1} & \frac{1}{s+0.1} \end{pmatrix} \tag{10}$$

that is to be controlled by applying an Internal Model Control (IMC) PI controller [13]. By using this IMC approach the desired closed loop dynamics and corresponding controller tuning are easily related by means of the IMC tuning parameter. Therefore, on one side we will have $T_i^d(s) = \frac{1}{\lambda_i s+1}$ and the corresponding controller tuning (assuming a First Order Plus Time delay representation for the plant): $K_p = \frac{T}{K(\lambda_i+L)}$ and $T_i = T$. From the previous discussion there will be a constraint on the allowable λ_i. For example, if we choose $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$, the gain (8) at the phase-crossover frequency is larger than one and the desired closed-loop gains have to be relaxed. Even the local loops are stable and meet the desired dynamics, the overall multivariable system becomes unstable due to the effect of interaction. If the performance specification for loop 1 is relaxed and the choice $\lambda_1 = 1.5$ is used, interaction generated on loop 2 is reduced as well as the total interaction in loop 1. Figure (2) shows the level of interacton that is generated in loop 2 has been reduced and the achieved closed loop relations (9) have the low-pass shape with some additional dynamics on loop1 introduced because of the closed-loop interaction.

Note the interaction level do not need to be below one for both loops simultaneously. As it can be confirmed by the stability constraints that are shown in figure (3) and the corresponding time responses plotted in figure (4). It can be appreciated that, effectively, loop 2 is affected by a low level of interaction than loop 1 (loop 2 still maintains a higher bandwith).

[1] even we are only dealing with nominal models the constraint apply as we see $G_i(s)$ as the result of modifying the $G_{ii}(s)$ dynamics

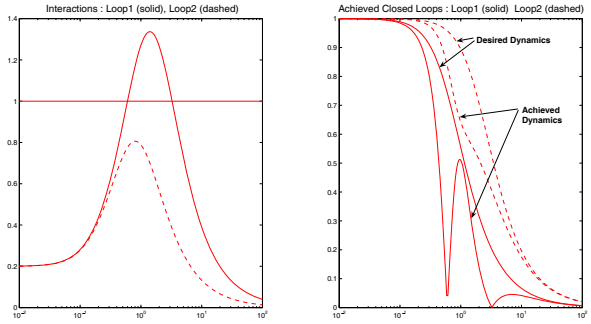


Figure 2: Interaction levels for both loops as well as the final achieved closed loop relations for $\lambda_1 = 1.5$

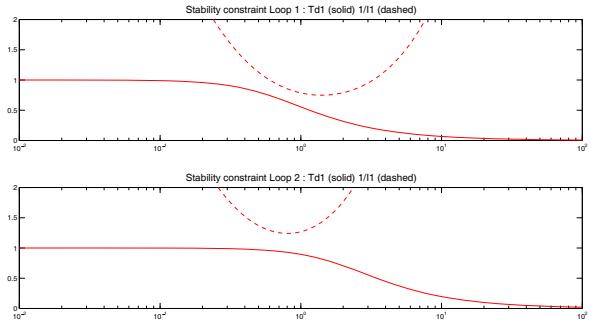


Figure 3: Stability Constraint for $\lambda_1 = 1.5$ and $\lambda_2 = 0.5$

5 Conclusions

A frequency domain analysis has been performed of the effects of interaction for decentralized control of a TITO process. Explicit use of the desired performance for each one of the independent loops has been used and introduced into the corresponding constraints for stability of the resulting multivariable closed loop. The introduction of an interaction measure associated to each loop in terms of the process interaction and desired dynamics appears to be a key term to take into account in order to test the stability as well as he achievable performance of the final closed loop with respect to the originally specified dynamics. Research us conducted on extension of this analysis to systems with more input and output signals as well as to obtain design (auto)tuning rules when controllers of PI/PID type are to be used.

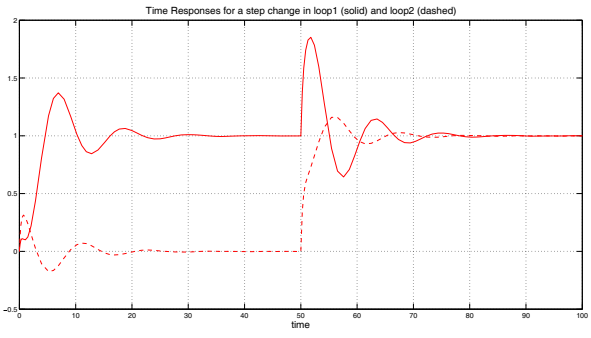


Figure 4: Step responses for a step change in both loops.

References

[1] E.F. Camacho and C. Bordons, *Model Predictive Control in the Process Industry*, Springer-Verlag, 1995.

[2] F.G. Shinskey, *Process Control Systems*, McGraw-Hill, 1996.

[3] E. Bristol, "On a new measure of interaction for multivariable process control," *IEEE Trans. Autom. Control*, vol. 11, 1966.

[4] A. Niederlinski, "A heuristic approach to the design of linear multivariable interacting subsystems," *Automatica*, vol. 7, pp. 691–701, 1971.

[5] M.F. Witcher and T.J. McAvoy, "Interacting control systems: steady-state and dynamic measurement of interaction," *ISA Trans*, vol. 16, pp. 35–41, 1977.

[6] E.H. Bristol, "Recent results on interaction in multivariable process control," *AIChE Conf. Miami*, 1978.

[7] J.E. Rijinsdrop, "Interaction in two-variable control systems for distillation," *Automatica*, vol. 1, pp. 15, 1965.

[8] M.S. Chiu and Y. Arkun, "A methodology for sequential design of robust decentralized control systems," *Automatica*, vol. 28, pp. 997–1001, 1992.

[9] M. Hovd and S. Skogestad, "Sequential design of decentralized controllers," *Automatica*, vol. 30, pp. 1601–1607, 1994.

[10] S. Skogestad and M. Morari, "Robust performance of decentralized control systems by independent designs," *Automatica*, vol. 25, pp. 119–125, 1989.

[11] M. Morari and E. Zafirou, *Robust Process Control.*, Prentice-Hall International, 1989.

[12] Z. Zhu and M. Chiu, "Dynamic analysis of decentralized 2x2 control systems in relation to loop interaction and local stability," *Ind. Eng. Chem. Res.*, vol. 37, pp. 464–473, 1998.

[13] R. Vilanova, "Revisiting imc based design of pi/pid controllers for foptd models," *11th IEEE International Conference on Emergin Technologies and Factory Automation, Prague, 20-22, September*, 2006.

Ramon Vilanova
Universitat Autonoma de Barcelona
Department of Telecomunications and Systems Engineering
ETSE, Campus UAB, 08193, Bellaterra, Barcelona
E-mail: Ramon.Vilanova@uab.cat

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 531-535

Some Properties of the Regular Asynchronous Systems

Şerban E. Vlad

Abstract: The asynchronous systems are the models of the asynchronous circuits from the digital electrical engineering. An asynchronous system f is a multi-valued function that assigns to each admissible input $u : \mathbf{R} \to \{0,1\}^m$ a set $f(u)$ of possible states $x \in f(u), x : \mathbf{R} \to \{0,1\}^n$. A special case of asynchronous system consists in the existence of a Boolean function $\Upsilon : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^n$ such that $\forall u, \forall x \in f(u)$, a certain equation involving Υ is fulfilled. Then Υ is called the generator function of f (Moisil used the terminology of network function) and we say that f is generated by Υ. The systems that have a generator function are called regular.
Our purpose is to continue the study of the generation of the asynchronous systems that was started in [2], [3].
Keywords: asynchronous system, regularity, generator function

1 Preliminaries

Notation 1. *Let be the arbitrary set M. The following notation will be useful:* $P^*(M) = \{M' | M' \subset M, M' \neq \emptyset\}$.

Definition 2. The set $\mathbf{B} = \{0,1\}$, endowed with the order $0 \leq 1$ and with the usual laws $-, \cdot, \cup, \oplus$, is called the **binary Boole algebra**.

Definition 3. The **initial value** $x(-\infty + 0) \in \mathbf{B}^n$ of the function $x : \mathbf{R} \to \mathbf{B}^n$ is defined by $\exists t' \in \mathbf{R}, \forall t < t', x(t) = x(-\infty + 0)$.

Definition 4. The **characteristic function** $\chi_A : \mathbf{R} \to \mathbf{B}$ of the set $A \subset \mathbf{R}$ is given by $\forall t \in \mathbf{R}, \chi_A(t) = \begin{cases} 1, t \in A \\ 0, \ else \end{cases}$.

Notation 5. *We use the notation* $Seq = \{(t_k) | t_k \in \mathbf{R}, k \in \mathbf{N}, t_0 < ... < t_k < ...$ *is unbounded from above*$\}$.

Definition 6. A function $x : \mathbf{R} \to \mathbf{B}^n$ is called $n-$**signal**, shortly **signal** if $\mu \in B^n$ and $(t_k) \in Seq$ exist such that

$$x(t) = \mu \cdot \chi_{(-\infty, t_0)}(t) \oplus x(t_0) \cdot \chi_{[t_0, t_1)}(t) \oplus ... \oplus x(t_k) \cdot \chi_{[t_k, t_{k+1})}(t) \oplus ... \tag{1}$$

The set of the $n-$signals is denoted by $S^{(n)}$.

Remark 7. Let be $x : \mathbf{R} \to \mathbf{B}^n, u : \mathbf{R} \to \mathbf{B}^m$. Instead of $x \times u : \mathbf{R} \times \mathbf{R} \to \mathbf{B}^n \times \mathbf{B}^m$ we define the function $x \times u$, many times denoted by (x,u), as $x \times u : \mathbf{R} \to \mathbf{B}^n \times \mathbf{B}^m$ due to the existence of a unique time variable $t \in \mathbf{R}$. Between the consequences derived from here we have the identifications $S^{(n)} \times S^{(m)} = S^{(n+m)}$ and $P^*(S^{(n)}) \times P^*(S^{(m)}) = P^*(S^{(n+m)})$.

Definition 8. The **left limit** $x(t-0)$ of $x(t)$ from (1) is the $\mathbf{R} \to \mathbf{B}^n$ function defined as $x(t-0) = \mu \cdot \chi_{(-\infty, t_0]}(t) \oplus x(t_0) \cdot \chi_{(t_0, t_1]}(t) \oplus ... \oplus x(t_k) \cdot \chi_{(t_k, t_{k+1}]}(t) \oplus ...$

Definition 9. Let be $U \in P^*(S^{(m)})$. A multi-valued function $f : U \to P^*(S^{(n)})$ is called **asynchronous system**, shortly **system**. Any $u \in U$ is called (**admissible**) **input** and the functions $x \in f(u)$ are called (**possible**) **states**.

Remark 10. The asynchronous systems are the models of the asynchronous circuits. The multi-valued character of the cause-effect association is due to the statistical fluctuations in the fabrication process, the variations in the ambiental temperature, the power supply etc. Sometimes the systems are given by equations and/or inequalities.

Definition 11. The **initial state function** of f is by definition the function $i_f : U \to P^*(\mathbf{B}^n), \forall u \in U, i_f(u) = \{x(-\infty + 0) | x \in f(u)\}$.

Definition 12. The function $\rho : \mathbf{R} \to \mathbf{B}^n$ is called **progressive** if $(t_k) \in Seq$ exists such that $\rho(t) = \rho(t_0) \cdot \chi_{\{t_0\}}(t) \oplus ... \oplus \rho(t_k) \cdot \chi_{\{t_k\}}(t) \oplus ...$ and $\forall i \in \{1, ..., n\}$, the set $\{k | k \in \mathbf{N}, \rho_i(t_k) = 1\}$ is infinite. The set of the progressive functions is denoted by P_n.

Notation 13. *Let be* $\Upsilon : \mathbf{B}^n \times \mathbf{B}^m \to \mathbf{B}^n, u \in S^{(m)}$, $\mu \in \mathbf{B}^n$ *and* $\rho \in P_n$. *The solution of the equation*

$$\left\{ \begin{array}{c} x(-\infty+0) = \mu \\ \forall i \in \{1,...,n\}, x_i(t) = \left\{ \begin{array}{c} \Upsilon_i(x(t-0), u(t-0)), if\ \rho_i(t) = 1 \\ x_i(t-0), otherwise \end{array} \right. \end{array} \right. \tag{2}$$

is denoted by $\Upsilon^{-\rho}(t, \mu, u)$.

Definition 14. The system $\Sigma_\Upsilon^- : S^{(m)} \to P^*(S^{(n)}), \forall u \in S^{(m)}, \Sigma_\Upsilon^-(u) = \{\Upsilon^{-\rho}(t, \mu, u) | \mu \in \mathbf{B}^n, \rho \in P_n\}$ is called the **universal regular asynchronous system** that is generated by the function Υ.

Definition 15. The system f is called **regular** if Υ exists such that $\forall u \in U, f(u) \subset \Sigma_\Upsilon^-(u)$. If so, Υ is called the **generator function** of f and we also say that Υ **generates** f.

Remark 16. Equation (2) shows how the circuits compute asynchronously the Boolean function Υ : the computation is made at the discrete time instances $\{t_k | k \in \mathbf{N}, \exists i \in \{1,...,n\}, \rho_i(t_k) = 1\}$ on these coordinates Υ_i for which $\rho_i(t_k) = 1$. The models of these circuits, the systems f with the generator function Υ, have the remarkable property that a function $\pi_f : W_f \to P^*(P_n)$ exists, $W_f = \{(x(-\infty+0), u) | u \in U, x \in f(u)\}$ such that $\forall u \in U, f(u) = \{\Upsilon^{-\rho}(t, \mu, u) | \mu \in i_f(u), \rho \in \pi_f(\mu, u)\}$. π_f is called the **computation function** of f. For f regular, Υ and π_f are not unique.

2 Subsystems

Definition 17. The system f is called a **subsystem** of $g : V \to P^*(S^{(n)}), V \in P^*(S^{(m)})$ and we write $f \subset g$, if $U \subset V$ and $\forall u \in U, f(u) \subset g(u)$.

Remark 18. We interpret $f \subset g$ in the following way: the systems f and g model the same circuit, but the model represented by f is more precise than the model represented by g.

Theorem 19. *The function* Υ *and the regular systems* $f \subset \Sigma_\Upsilon^-$, $g \subset \Sigma_\Upsilon^-$ *are given. We denote by* $i_g : V \to P^*(\mathbf{B}^n)$ *the initial state function and by* $\pi_g : W_g \to P^*(P_n)$ *the computation function of* g. *The following statements are equivalent:*

a) $f \subset g$

b) $U \subset V$ *and* $\forall u \in U, i_f(u) \subset i_g(u)$ *and* $\forall u \in U, \forall \mu \in i_f(u), \forall \rho \in \pi_f(\mu, u), \exists \rho' \in \pi_g(\mu, u), \Upsilon^{-\rho}(t, \mu, u) = \Upsilon^{-\rho'}(t, \mu, u)$.

3 Dual systems

Definition 20. The **dual** function $\Upsilon^* : \mathbf{B}^n \times \mathbf{B}^m \to \mathbf{B}^n$ of Υ is defined by $\forall (\mu, \nu) \in \mathbf{B}^n \times \mathbf{B}^m, \Upsilon^*(\mu, \nu) = \overline{\Upsilon(\overline{\mu}, \overline{\nu})}$. Here the bar $\overline{\mu}$ refers to the complement done coordinatewise.

Definition 21. The **dual** of the system f is by definition the system $f^* : U^* \to P^*(S^{(n)})$, where $U^* = \{\overline{u} | u \in U\}$ and $\forall u \in U^*, f^*(u) = \{\overline{x} | x \in f(\overline{u})\}$.

Remark 22. The system f^* models the circuit modeled by f with the AND gates replaced by OR gates etc.

Notation 23. *We denote* $i_{f^*} : U^* \to P^*(\mathbf{B}^n)$, $\forall u \in U^*, i_{f^*}(u) = \{\overline{\mu} | \mu \in i_f(\overline{u})\}$.

Notation 24. *We denote by* $\pi_{f^*} : W_{f^*} \to P^*(P_n)$ *where* $W_{f^*} = \{(\overline{x(-\infty+0)}, u) | u \in U^*, x \in f(\overline{u})\}$ *the function* $\forall (\mu, u) \in W_{f^*}, \pi_{f^*}(\mu, u) = \pi_f(\overline{\mu}, \overline{u})$.

Theorem 25. *The dual system* f^* *of* $f \subset \Sigma_\Upsilon^-$ *is regular,* $f^* \subset \Sigma_{\Upsilon^*}^-$; *its initial state function is* i_{f^*} *and its computation function is* π_{f^*}.

4 Cartesian product

Definition 26. The **Cartesian product** of the systems f and $f' : U' \to P^*(S^{(n')}), U' \in P^*(S^{(m')})$ is defined as $f \times f' : U \times U' \to P^*(S^{(n+n')}), \forall (u,u') \in U \times U', (f \times f')(u,u') = f(u) \times f'(u')$.

Remark 27. The Cartesian product $f \times f'$ models two circuits that run independently on each other.

Notation 28. *For Υ and $\Upsilon' : \mathbf{B}^{n'} \times \mathbf{B}^{m'} \to \mathbf{B}^{n'}$, we denote by $\Upsilon \times \Upsilon' : \mathbf{B}^{n+n'} \times \mathbf{B}^{m+m'} \to \mathbf{B}^{n+n'}$ the function $\forall ((\mu,\mu'),(\nu,\nu')) \in \mathbf{B}^{n+n'} \times \mathbf{B}^{m+m'}, (\Upsilon \times \Upsilon')((\mu,\mu'),(\nu,\nu')) = (\Upsilon(\mu,\nu),\Upsilon'(\mu',\nu'))$. In this notation we identify $(\mu,\mu') \in \mathbf{B}^n \times \mathbf{B}^{n'}$ with $(\mu_1,...,\mu_n,\mu'_1,...,\mu'_{n'}) \in \mathbf{B}^{n+n'}$ etc.*

Notation 29. *If $i_{f'} : U' \to P^*(\mathbf{B}^{n'})$ is the initial state function of f', we use the notation $i_{f \times f'} : U \times U' \to P^*(\mathbf{B}^{n+n'}), \forall (u,u') \in U \times U', i_{f \times f'}(u,u') = i_f(u) \times i_{f'}(u')$.*

Notation 30. *The regular systems f, f' are given, $f \subset \Sigma_\Upsilon$, $f' \subset \Sigma_{\Upsilon'}$ as well as their computation functions: $\pi_f : W_f \to P^*(P_n)$, $\pi_{f'} : W_{f'} \to P^*(P_{n'})$. We denote by $\pi_{f \times f'} : W_{f \times f'} \to P^*(P_{n+n'})$ the function $W_{f \times f'} = \{((x(-\infty+0),x'(-\infty+0)),(u,u'))|(u,u') \in U \times U', (x,x') \in f(u) \times f'(u')\}, \forall ((\mu,\mu'),(u,u')) \in W_{f \times f'}, \pi_{f \times f'}((\mu,\mu'),(u,u')) = \pi_f(\mu,u) \times \pi_{f'}(\mu',u')$.*

Remark 31. The function $\pi_{f \times f'}$ is correctly defined since $\forall \rho, \forall \rho', \rho \in P_n$ and $\rho' \in P_{n'} \Longrightarrow (\rho,\rho') \in P_{n+n'}$.

Theorem 32. *If $f \subset \Sigma^-_\Upsilon, f' \subset \Sigma^-_{\Upsilon'}$, then the system $f \times f'$ is regular, $f \times f' \subset \Sigma^-_{\Upsilon \times \Upsilon'}$; its initial state function is $i_{f \times f'}$ and its computation function is $\pi_{f \times f'}$.*

5 Parallel connection

Definition 33. The **parallel connection** of f and $f'_1 : U'_1 \to P^*(S^{(n')}), U'_1 \in P^*(S^{(m)})$ is defined whenever $U \cap U'_1 \neq \emptyset$ by $f||f'_1 : U \cap U'_1 \to P^*(S^{(n+n')}), \forall u \in U \cap U'_1, (f||f'_1)(u) = f(u) \times f'_1(u)$.

Remark 34. The parallel connection $f||f'_1$ models two circuits that run under the same input, independently on each other.

Notation 35. *Let be Υ and $\Upsilon'_1 : \mathbf{B}^{n'} \times \mathbf{B}^m \to \mathbf{B}^{n'}$, for which we denote by $\Upsilon||\Upsilon'_1 : \mathbf{B}^{n+n'} \times \mathbf{B}^m \to \mathbf{B}^{n+n'}$ the function $\forall ((\mu,\mu'),\nu) \in \mathbf{B}^{n+n'} \times \mathbf{B}^m, (\Upsilon||\Upsilon'_1)((\mu,\mu'),\nu) = (\Upsilon(\mu,\nu),\Upsilon'_1(\mu',\nu))$.*

Notation 36. *Let $i_{f'_1} : U'_1 \to P^*(\mathbf{B}^{n'})$ be the initial state function of f'_1. If $U \cap U'_1 \neq \emptyset$, we use the notation $i_{f||f'_1} : U \cap U'_1 \to P^*(\mathbf{B}^{n+n'}), \forall u \in U \cap U'_1, i_{f||f'_1}(u) = i_f(u) \times i_{f'_1}(u)$.*

Notation 37. *We suppose that the systems f, f'_1 are regular i.e. $f \subset \Sigma^-_\Upsilon, f'_1 \subset \Sigma^-_{\Upsilon'_1}$ and let $\pi_f : W_f \to P^*(P_n), \pi_{f'_1} : W_{f'_1} \to P^*(P_{n'})$ be their computation functions. If $U \cap U'_1 \neq \emptyset$, then we use the notation $\pi_{f||f'_1} : W_{f||f'_1} \to P^*(P_{n+n'})$, $W_{f||f'_1} = \{((x(-\infty+0),x'(-\infty+0)),u)|u \in U \cap U'_1, x \in f(u), x' \in f'_1(u)\}, \forall ((\mu,\mu'),u) \in W_{f||f'_1}, \pi_{f||f'_1}((\mu,\mu'),u) = \pi_f(\mu,u) \times \pi_{f'_1}(\mu',u)$.*

Theorem 38. *If $f \subset \Sigma^-_\Upsilon$, $f'_1 \subset \Sigma^-_{\Upsilon'_1}$ and $U \cap U'_1 \neq \emptyset$, then $f||f'_1 \subset \Sigma^-_{\Upsilon||\Upsilon'_1}$; its initial state function is $i_{f||f'_1}$ and its computation function is $\pi_{f||f'_1}$.*

6 Serial connection

Remark 39. Let be the systems f and $h : X \to P^*(S^{(p)}), X \in P^*(S^{(n)})$. When $\bigcup\limits_{u \in U} f(u) \subset X$, the serial connection of f and h is defined by $h \circ f : U \to P^*(S^{(p)}), \forall u \in U, (h \circ f)(u) = \bigcup\limits_{x \in f(u)} h(x)$. If f and h are regular, this definition means that in the systems of equations

$$\left\{ \begin{array}{c} x(-\infty+0) = \mu \\ \forall i \in \{1,...,n\}, x_i(t) = \left\{ \begin{array}{c} \Upsilon_i(x(t-0),u(t-0)), if\ \rho_i(t) = 1 \\ x_i(t-0), otherwise \end{array} \right. \end{array} \right. , \tag{3}$$

$$\left\{\begin{array}{c} y(-\infty+0)=\lambda \\ \forall j\in\{1,...,p\}, y_j(t)=\left\{\begin{array}{c} \vartheta_j(y(t-0),x(t-0)), if\ \varpi_j(t)=1 \\ y_j(t-0), otherwise \end{array}\right. \end{array}\right. \tag{4}$$

where $u\in S^{(m)}, x\in S^{(n)}, y\in S^{(p)}, \mu\in \mathbf{B}^n,\ \lambda\in\mathbf{B}^p,\ \rho\in P_n,\ \varpi\in P_p,\ \Upsilon:\mathbf{B}^n\times\mathbf{B}^m\to\mathbf{B}^n,\ \vartheta:\mathbf{B}^p\times\mathbf{B}^n\to\mathbf{B}^p$ we eliminate x. Because this does not give any information of the regularity of $h\circ f$, we choose to work with a slightly different system from $h\circ f$, for which x is not eliminated.

Notation 40. *If f and h fulfill $\bigcup_{u\in U} f(u)\subset X$, then we denote by $h*f:U\to P^*(S^{(n+p)})$ the system $\forall u\in U, (h*f)(u)=\{(x,y)|x\in f(u), y\in h(x)\}$.*

Notation 41. *The function $\vartheta*\Upsilon:\mathbf{B}^{n+p}\times\mathbf{B}^m\to\mathbf{B}^{n+p}$ is defined by $\forall((\mu,\lambda),\nu)\in\mathbf{B}^{n+p}\times\mathbf{B}^m, (\vartheta*\Upsilon)((\mu,\lambda),\nu)=(\Upsilon(\mu,\nu),\vartheta(\lambda,\Upsilon(\mu,\nu)))$.*

Remark 42. The point is that, instead of eliminating x in (3), (4) as $h\circ f$ does, we can work with $h*f$ and with the equation

$$\left\{\begin{array}{c} z(-\infty+0)=(\mu,\lambda) \\ \forall k\in\{1,...,n+p\}, z_k(t)=\left\{\begin{array}{c} (\vartheta*\Upsilon)_k(z(t-0),u(t-0)), if\ (\rho,\varpi)_k(t)=1 \\ z_k(t-0), otherwise \end{array}\right. \end{array}\right.$$

where $z\in S^{(n+p)}$.

Notation 43. *For $i_h:X\to P^*(\mathbf{B}^p)$ the initial state function of h, we denote by $i_{h*f}:U\to P^*(\mathbf{B}^{n+p})$ the function $\forall u\in U, i_{h*f}(u)=\{(\mu,\lambda)|\mu\in i_f(u), \lambda\in\bigcup_{x\in f(u), x(-\infty+0)=\mu} i_h(x)\}$.*

Notation 44. *We suppose that $\pi_h:W_h\to P^*(P_p)$ is the computation function of h, $W_h=\{(y(-\infty+0),x)|x\in X, y\in h(x)\}$. We denote by $\pi_{h*f}:W_{h*f}\to P^*(P_{n+p})$ the function $W_{h*f}=\{((x(-\infty+0),y(-\infty+0)),u)|u\in U, x\in f(u), y\in h(x)\}$, $\forall((\mu,\lambda),u)\in W_{h*f}$, $\pi_{h*f}((\mu,\lambda),u)=$*
$\{(\rho,\varpi)|\rho\in\pi_f(\mu,u), \varpi\in\bigcup_{x\in f(u), x(-\infty+0)=\mu}\pi_h(\lambda,x)\}$.

Theorem 45. *The systems f and h are given such that the inclusion $\bigcup_{u\in U} f(u)\subset X$ is true. If the regularity properties $f\subset\Sigma^-_\Upsilon, h\subset\Sigma^-_\vartheta$ hold, then $h*f\subset\Sigma^-_{\vartheta*\Upsilon}$; the initial state function of $h*f$ is i_{h*f} and its computation function is π_{h*f}.*

7 Intersection

Definition 46. The **intersection** of $f:U\to P^*(S^{(n)})$ and $g:V\to P^*(S^{(n)}), U,V\in P^*(S^{(m)})$ is defined whenever $\exists u\in U\cap V, f(u)\cap g(u)\neq\emptyset$ by $f\cap g:W\to P^*(S^{(n)})$, $W=\{u|u\in U\cap V, f(u)\cap g(u)\neq\emptyset\}$, $\forall u\in W, (f\cap g)(u)=f(u)\cap g(u)$.

Remark 47. The intersection of two systems is a model that results by the simultaneous validity of two compatible models.

Notation 48. *When $W\neq\emptyset$, we use the notation $i_{f\cap g}:W\to P^*(\mathbf{B}^n), \forall u\in W, i_{f\cap g}(u)=i_f(u)\cap i_g(u)$.*

Notation 49. *We consider the regular systems f,g for which the generator function $\Upsilon:\mathbf{B}^n\times\mathbf{B}^m\to\mathbf{B}^n$ is given such that $f\subset\Sigma^-_\Upsilon, g\subset\Sigma^-_\Upsilon$. Their computation functions are $\pi_f:W_f\to P^*(P_n)$, $\pi_g:W_g\to P^*(P_n)$. If the set W is non-empty, then we use the notation $\pi_{f\cap g}:W_{f\cap g}\to P^*(P_n)$ for the function that is defined by $W_{f\cap g}=\{(x(-\infty+0),u)|u\in W, x\in f(u)\cap g(u)\}, \forall(\mu,u)\in W_{f\cap g}$, $\pi_{f\cap g}(\mu,u)=\{\rho|\rho\in\pi_f(\mu,u), \exists\rho'\in\pi_g(\mu,u), \Upsilon^{-\rho}(t,\mu,u)=\Upsilon^{-\rho'}(t,\mu,u)\}$.*

Remark 50. We remark the satisfaction of the following property of symmetry: $W_{f\cap g}=W_{g\cap f}$ and $\forall(\mu,u)\in W_{f\cap g}, \forall\rho\in\pi_{f\cap g}(\mu,u), \exists\rho'\in\pi_{g\cap f}(\mu,u), \Upsilon^{-\rho}(t,\mu,u)=\Upsilon^{-\rho'}(t,\mu,u)$ and $\forall\rho'\in\pi_{g\cap f}(\mu,u), \exists\rho\in\pi_{f\cap g}(\mu,u), \Upsilon^{-\rho'}(t,\mu,u)=\Upsilon^{-\rho}(t,\mu,u)$.

Theorem 51. *If the regular systems $f\subset\Sigma^-_\Upsilon, g\subset\Sigma^-_\Upsilon$ fulfill $W\neq\emptyset$, then their intersection $f\cap g:W\to P^*(S^{(n)})$ is regular $f\cap g\subset\Sigma^-_\Upsilon$; its initial state function is $i_{f\cap g}$ and its computation function is $\pi_{f\cap g}$.*

8 Union

Definition 52. The **union** of f,g is defined by $f\cup g:U\cup V\to P^*(S^{(n)}),\forall u\in U\cup V,(f\cup g)(u)=\begin{cases} f(u),u\in U\setminus V,\\ g(u),u\in V\setminus U,\\ f(u)\cup g(u),u\in U\cap V\end{cases}$.

Remark 53. The union of the systems represents the validity of one of two models. This is useful for example in testing, when f is the model of the 'good' circuit and g is the model of the 'bad' circuit.

Notation 54. *We denote by* $i_{f\cup g}:U\cup V\to P^*(\mathbf{B}^n)$ *the function* $\forall u\in U\cup V,i_{f\cup g}(u)=\begin{cases} i_f(u),u\in U\setminus V,\\ i_g(u),u\in V\setminus U,\\ i_f(u)\cup i_g(u),u\in U\cap V\end{cases}$.

Lemma 55. *The sets* $W_f=\{(x(-\infty+0),u)|u\in U,x\in f(u)\}$, $W_g=\{(x(-\infty+0),u)|u\in V,x\in g(u)\}$, $W_{f\cup g}=\{(x(-\infty+0),u)|u\in U\cup V,x\in(f\cup g)(u)\}$ *fulfill* $W_{f\cup g}=W_f\cup W_g$.

Notation 56. *Let be the regular systems* f, g *and the function* Υ *such that* $f\subset\Sigma_\Upsilon^-,g\subset\Sigma_\Upsilon^-$ *are true. The computation functions of* f,g *are* π_f, π_g. *We denote by* $\pi_{f\cup g}:W_{f\cup g}\to P^*(P_n)$ *the function* $\forall(\mu,u)\in W_{f\cup g}$, $\pi_{f\cup g}(\mu,u)=\begin{cases}\pi_f(\mu,u),(\mu,u)\in W_f\setminus W_g,\\ \pi_g(\mu,u),(\mu,u)\in W_g\setminus W_f,\\ \pi_f(\mu,u)\cup\pi_g(\mu,u),(\mu,u)\in W_f\cap W_g\end{cases}$.

Theorem 57. *If the systems* f,g *are regular* $f\subset\Sigma_\Upsilon^-,g\subset\Sigma_\Upsilon^-$, *then the union* $f\cup g:U\cup V\to P^*(S^{(n)})$ *is regular,* $f\cup g\subset\Sigma_\Upsilon^-$; *its initial state function is* $i_{f\cup g}$ *and its computation function is* $\pi_{f\cup g}$.

References

[1] S. E. Vlad, *Teoria sistemelor asincrone,* Editura Pamantul Pitesti and WSEAS Press Athens, 2007.

[2] S. E. Vlad, "Boolean dynamical systems", *The 15-th Conference on Applied and Industrial Mathematics CAIM 2007, Mioveni, Romania, October 12-14*, 2007.

[3] S. E. Vlad, "On the generation of the asynchronous systems", *The first National Conference of Applied and Fundamental Mathematics, Iasi, Romania, November 9-10*, 2007.

Şerban E. Vlad
Oradea City Hall
Computers Department
Piaţa Unirii, Nr. 1, 410100, Oradea, Romania
E-mail: serban_e_vlad@yahoo.com

Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844
Vol. III (2008), Suppl. issue: Proceedings of ICCCC 2008, pp. 536-540

Issues on Optimality Criteria Applied in Real-Time Scheduling

Doina Zmaranda, Gianina Gabor

Abstract: Task scheduling is a wide research area with several scheduling techniques, each of which being based on a specific algorithmic approach. Although these techniques are often adjusted to the specifics of the various scheduling problems, there are few that are optimal for real-time domain. The aim of this paper is to give an overview the main algorithmic approaches and optimality criteria that could be applied in real time scheduling. Different methods for solving selected problem classes are discussed.
Keywords:time scheduling, schedule length, lateness, tardiness, flow time

1 Introduction

In several real-time applications tasks are not only subjected to complete by their deadlines, but more control of the schedule is required. Consequently, quality of service could be interpreted in many ways: jitters, response times, and other. Thus, optimization of a performance measure while respecting release times and deadlines of the tasks is an important issue.
In non real-time environments, response time or flow time is a widely used performance metric to optimize continuous jobs arrivals. When each task has to be responsive in a single processor problem, then the optimized performance measure is the maximum response time of the tasks. If the whole system has to be responsive[11], then the optimized metric could be considered the average response time, generally the shortest remaining time rule leading to an optimal schedule. Also, weighted versions of these problems has also been studied: for example weighted sum of response times, so called mean weighted flow time, when usually weights are the inverse of the processing times. The computational complexity of minimizing the mean weighted flow time is not known [2]; but, if arbitrary weights are allowed, then minimizing the weighted total response times is NP-hard in the strong sense [7].
Scheduling problems can be viewed as the problems of allocation of resources over time to perform a set of tasks. In the case of real-time scheduling, besides the other criteria, the usefulness of computational result depends also on the time limit it is produced. Also, many today's application areas implies complex and distributed systems to control and coordinate time critical activities. In real-time systems, an important concern in the analysis and development of scheduling strategies is predictability of system's behavior. In manufacturing environments, deterministic or predictive scheduling is preferred: that assumes that is possible to predict temporal behavior of all tasks.
If there is no sufficient knowledge to predict system's behavior, sometimes reactive scheduling using a shorter planning horizon could lead to acceptable results. Unfortunately, in many situations, especially if deadlines have to be met, the reactive approach could not be applied. Also, when tasks processing times are unknown, the only way to solve this problem is to assume upper bounds to the processing times; consequently, if all deadlines are met with respect to these upper bounds, no deadline will be exceeded for the real task processing times. A broad class of real-time computer control systems uses this approach [10].
The purpose of task scheduling is to organize a set of tasks ready for execution by a processor system, i.e. to organize them so that performance objectives are met. Thus, scheduling is essentially an optimization problem. The order or arrangement of these tasks is called a schedule; in real-time systems, the primary criterion for a schedule is to ensure that all tasks meet their deadlines. A schedule can be feasible or optimal: a feasible schedule orders tasks so that all meet their deadlines; an optimal schedule is one which ensures that failures to meet tasks deadlines are minimized [8].

2 Applying optimality criteria in real-time scheduling

For real-time systems, task scheduling represents a complex optimization problem, most of the scheduling policies being limited to one or, at most, two constrains. The question that arises is if existing approaches could generate optimal task performance in real-time systems.In constructing a task scheduling model, the following items should be considered:

- a set of tasks, $T = \{T_1, T_2 \cdots T_n\}$
- a set of processors $P = \{P_1, P_2 \cdots P_m\}$
- a set of necessary resources $R = \{R_1, R_2 \cdots R_p\}$

A feasible scheduling scheme means a particular assignment of processors from P and resources from R to tasks from T, assignment that implies completion of all tasks under certain imposed constraints. Each task T_j is characterized by the following data [4]:

- vector of processing times: p_j=[p_{1j}, p_{2j}, ... p_{mj}] , where p_{ij} is he time needed by processor P_i to complete task T_j. If all processors are identical, p_{ij}=p_j for $i = 1 \cdots m$
- precedence constrains among tasks. $T_i < T_j$ means that processing of T_i must be completed before T_j can be started. That means that on the T set a partially ordered relation is defined
- release time or ready time r_j is the time at which task T_j is ready for execution
- the deadline or due time d_j for the task T_j - specifies the time limit before the task T_j should be completed
- periodicity of tasks: if periodic tasks are considered, task period for task T_j is c_j
- priority of tasks, that expresses the relative urgency of tasks, u_i

In deterministic scheduling theory, a priori knowledge of release and processing times is assumed. On the other way, as far as processing times are concerned, generally they are very difficult to be estimated in certain circumstances. This leads to a more flexible approach of using upper bounds on the processing times [5]. In hard real-time environments with given tasks deadlines this approach is often used; if all deadlines are met with respect to their upper bounds, no deadline will be exceeded for the real task processing times. Another possibility is to take into consideration the mean task processing times instead of exact values and calculate optimistic estimate of the mean value of the schedule length. Generally, the following parameters can be calculated for each task processed by a given schedule:

- completion time for a task T_i, denoted by C_i
- flow time for task T_i, denoted by F_i, being the sum of waiting and processing times:

$$F_i = C_i - r_i \tag{1}$$

- latenss, denoted by L_i for a task T_i,

$$L_i = C_i - d_i \tag{2}$$

- tardiness denoted by $Tard_i$,

$$Tard_i = max\{ci - di, 0\} \tag{3}$$

For a given schedule, there are known several optimality criteria that could be calculated in order to evaluate the schedule performance:

- schedule length L_{sch} where C_j is completion time for task T_j :

$$L_{sch} = max\{C_j\} \tag{4}$$

- mean flow time F where F_j is C_j-r_j :

$$F = \frac{1}{n} \sum_{j=1}^{n} Fj \tag{5}$$

- mean weighted flow time :

$$F_u = (\sum_{j=1}^{n} u_j * F_j)/(\sum_{j=1}^{n} u_j) \tag{6}$$

Minimizing schedule length leads to maximization of processor utilization within makespan L_{sch} and also to the minimization of the maximum in-process time of the scheduled set of tasks. The mean flow time minimization yields a minimization of the mean response time and the mean in process time of the scheduled set of tasks.
But, for real-time applications performance measures take into account other issues, such as lateness or tardiness of tasks, as they are defined in (2) and (3). Consequently, specific real-time optimality criteria could be some of the following ones [6]:

- maximum lateness :

$$L_{max} = max\{L_j\} \tag{7}$$

- the mean tardiness :

$$Tard_{med} = \frac{1}{n}\sum_{j=1}^{n} Tard_j \tag{8}$$

- the mean weighted tardiness :

$$T = (\sum_{j=1}^{n} u_j * Tard_j)/(\sum_{j=1}^{n} u_j) \tag{9}$$

- the number of tardy tasks N_t where $N_{tard\,j} = 1$ if $C_j \geq d_j$ and 0 otherwise :

$$N_t = \sum_{j=1}^{n} N_{tard\,j} \tag{10}$$

- weighted number of tardy tasks :

$$N_u = \sum_{j=1}^{n} u_j * Tard_j \tag{11}$$

These above deadline involving criteria are of great importance in many real-time applications including control or manufacturing systems, since their minimization leads to construction of schedulers with no late tasks whenever such schedule exists. Generally, a schedule for which the value of a particular performance measure is as its minimum will be considered optimal. The criteria mentioned above are basic in the sense that they require specific approaches to the construction of schedulers [3]. A scheduling algorithm is an algorithm that constructs a schedule for a given problem. In general, optimization algorithms are considered, but, because of the inherent complexity of them, also approximation or heuristic algorithms are applied.
Scheduling problem belong to the broad class of combinatorial search problems, these optimization problems being called NP-hard. For real-time, it is obvious that the time available for solving particular scheduling problems is seriously limited, thus only low order polynomial time algorithms can be applied in order to solve the problem optimally in the time bounded.
The criteria mentioned in the introduction part are basic, in the sense that they require specific approaches to the construction of schedulers. In general, there is an important research activity in optimization criteria, but because of the inherent complexity of many problems of that type, sometimes solving scheduling problems is seriously limited in practice.

2.1 Minimizing schedule length

Complexity analysis of this problem leads to the conclusion that an optimization polynomial time algorithm could be find only if some relaxation of initial conditions is done: for example by allowing task preemption. In this case, the length of a preemptive schedule cannot be smaller than the maximum of the following values: maximum processing time of a task and mean processing requirement on a processor:

$$Cmax \geq max\{max\{p_j\}, \frac{1}{m}\sum_{j=1}^{n} p_j\} \tag{12}$$

In order to minimize the problem complexity when applying in practice, an usual approximation strategy used is to construct a priority list of the given tasks, and each step the first available processor is selected to process the first available task of the list. Obviously, the accuracy of a given list scheduling algorithm depends on the order in which tasks appear in the list.
Unfortunately, this simple strategy could weaken the precedence constrains, unless the initial list is constructed with respect to topological order of the processes. The way in which the initial task list is constructed has significant impact on the resulting schedule length. Generally, an arbitrary lists scheduling can produce schedulers almost twice as long as optimal ones. An improvement could be gained if tasks are ordered properly in the list, for example, in order of non-increasing processing times p_j. In this particular case, starting from this pre-ordered set of n tasks that must be scheduled on m processors, a possible scheduling algorithm could be:

```
for i=1 to m do si=0; //si are idle times for the m processors
j=1;
do
    sk=min {si};  // assign task Tj to processor Pk at time sk;
    sk = sk +pj;  j++;
while j<=n;       // al tasks have been scheduled
```

The algorithm behaves well in practice, especially when the number of tasks becomes very large.

2.2 Minimizing flow time

Minimizing flow time implies to schedule a set of task in such a way that the weighted sum of completion times is minimal. The problem can be solved by scheduling the tasks in order of non-decreasing ratios of processing times an weights, p_j/u_j.
If the schedule is constructed by adding one task at a time, starting from an empty schedule, at any point we will have a partial schedule, until the process is finished. If the order of tasks execution is restricted by arbitrary precedence constrains, then analyzing conflicts between the non-decreasing rations and a possible topological order should be done. If it is possible to find a solution that satisfies both criteria, then the resulting schedule will also comply to precedence constrains, so it will be the optimal solution.
The problem of minimizing the sum of weighted completion times is NP-hard, even if all weights are assumed to be 1. Practically it could be solved using two heuristic algorithms for scheduling, each one specifying the priority criteria for adding a task to an existing partial schedule:

- *earliest completion time algorithm* - which select the task T_i with minimum completion time $C_i = s_i + p_i$, where s_i is the earliest start time of task T_i and p_i is the task processing time
- *earliest start time algorithm*- which select task T_i with minimum start time s_i

For both above algorithms, no accuracy bounds are known.

2.3 Minimizing the number of tardy tasks and/or mean weighted tardiness

This criterion implies the minimization of the weighted number of tasks exceeding their deadlines. The simplest way to achieve this is to use an optimization algorithm that first sorts the tasks according to the Earliest Due Deadline First (EDF) rule [9]. Then, the subset of tasks that are tardy could be evaluated; for the case of un-weighted tasks, applying this algorithm generates a schedule with minimum number of tardy tasks.
Also, when all processing times are equal, the problem could be similarly solved. For the more general case, with agreeable processing times and weights, and different releases times, optimal schedules can be constructed using a simple modification in EDF algorithm, but only if all release and due times are consistent: $r_i < r_j$ implies $d_i < d_j$ for all tasks pairs T_i,T_j.
Another possibility of evaluating schedulers is deadline which is involved in mean tardiness. Generally, the mean weighted tardiness problem is NP-hard, even if all weights are equal. But if, in addition to this, if the assumption that all tasks are independent and have unit processing times still hold, the mean weighted tardiness can be minimized by simply sequencing tasks in non-decreasing order of their deadline and problem can be thus easily solved.

2.4 Scheduling with deadlines

In this case, the schedule can be constructed around the earliest-deadline-first principle. This simple algorithm produces optimal schedulers for real-time systems only under very restricted assumptions. The preemptive mode of processing makes the solution of scheduling problem much more easier.
A general approach for this case is to assign modified deadlines, depending on the number and successors of tasks, in order to respect the imposed precedence constraints. Also, there are other issues that may influence scheduling with deadlines: for example, when the start time and the deadline for a task from a periodic task set do not coincide, there is difficult problem to decide if such a periodic task set can be scheduled using earliest deadline first algorithm. The same difficulty arises when semaphores and mutual exclusion are used.
Another possibility is to use a heuristic algorithm that chooses one task of largest processing time among all tasks, and schedules this task last. Afterwards, it continues by choosing the largest processing task from the n-1 remaining tasks and so on. The algorithm is usable under the assumptions that no precedence constraints are imposed.

3 Summary and Conclusions

Task scheduling is a wide research area, with a great variety of algorithmic methods. But, for specific domains, as for example real-time domain, these methods must be adjusted and approximated to the specifics of the scheduling problems according to these fields. In this paper, some of the most important algorithmic approaches that could be applied to real-time scheduling are investigated.

References

[1] S. K. Baruah, L.E. Rosier, R. R. Howell,"Algorithms and Complexity Concerning the Preemptive Scheduling of Periodic, Real-Time Tasks on one Processor," *Real-Time Systems*, no.2, Kluwer Academic Publishers pp.301-324, 1990.

[2] C. Chejuri, S. Khanna, A. Zhu,"Algorithms for minimizing weighted flow time,"*Proc. ACM Symp. on Theory of Computing*, 2001.

[3] I. Bate, A. Burns,"A Framework for Scheduling in Safety-Critical Embedded Control Systems,"*Proceedings of the 6th International Conference on Real-Time Computing Systems and Applications*, pp.467-475, 1999.

[4] P. Kaminsky, D. Simchi-Levi,"A Framework for Scheduling in Safety-Critical Embedded Control Systems,"*Operations Research Leters*, no.29, pp.141-148, 2001.

[5] B. Korousic-Seljac,"Task Scheduling Policies for Real Time Systems,"*Microprocessors and Microsystems* ,vol.18, no.9, pp.501-511, 1994.

[6] J. Liu,*Real Time Systems,* Prentice Hall, 2000.

[7] P. Richard,"A Tool for Controlling Response Time in Real-Time Systems", TOOLS 2002, LNCS 2324, pp. 339-348, Springer-Verlag 2002.

[8] M. Spuri, J. A. Stankovic,"How to Integrate Precedence Constrains and Shared Resources in real-Time Scheduling,"*IEEE Transanctions on Computers*,vol 43, no. 12, pp. 1407-1412, 1994.

[9] J. Stankovic, K. Ramamritham,"Deadline Scheduling for Real-time systems: EDF and Related Algorithm,"*Kluwer Academic Publishers*, 1998.

[10] J. Stankovic,M. Spuri, M. Di Natale, and G. Buttazzo,"Implications of Classical Scheduling Results For Real-Time Systems,"*IEEE Transactions on Computers*, vol.28, no.6, 1995.

[11] J.K. Strosnider, J.P. Lehoczky, Sha Lui,,"The Deferrable Server Algorithm for Enhanced Aperiodic Responsiveness in Hard Real-Time Environments,"*IEEE Transactions on Computers*, vol.44, no.1, 1995.

Doina Zmaranda, Gianina Gabor
University of Oradea
Departmentof Computer Science
1 Universitatii St., Oradea, Romania
E-mail: {zdoina,gianina}@uoradea.ro

Author index

Description

International Journal of Computers, Communications & Control (IJCCC) is a quarterly peer-reviewed publication started in 2006 by Agora University Editing House - CCC Publications, Oradea, ROMANIA.

Beginning with 2007, EBSCO Publishing is a licensed partner of IJCCC Publisher.

Every issue is published in online format (ISSN 1841-9844) and print format (ISSN 1841-9836).

Now we offer free online access to the full text of all published papers.

The printed version of the journal should be ordered, by subscription, and will be delivered by regular mail.

IJCCC is directed to the international communities of scientific researchers from the universities, research units and industry.

IJCCC publishes original and recent scientific contributions in the following fields:

- Computing & Computational Mathematics
- Information Technology & Communications
- Computer-based Control

To differentiate from other similar journals, the editorial policy of IJCCC encourages especially the publishing of scientific papers that focus on the convergence of the 3 "C" (Computing, Communication, Control).

The articles submitted to IJCCC must be original and previously unpublished in other journals. The submissions will be revised independently by minimum two reviewers and will be published only after end of the editorial workflow.

The peer-review process is single blinded: the reviewers know who the authors of the manuscript are, but the authors do not have access to the information of who the peer-reviewers are.

IJCCC also publishes:

- papers dedicated to the works and life of some remarkable personalities;
- reviews of some recent important published books

Also, IJCCC will publish as supplementary issues the proceedings of some international conferences or symposiums on Computers, Communications and Control, scientific events that have reviewers and program committee.

The authors are kindly asked to observe the rules for typesetting and submitting described in Instructions for Authors.

Editorial Workflow

The editorial workflow is performed using the online Submission System.

The peer-review process is single blinded: the reviewers know who the authors of the manuscript are, but the authors do not have access to the information of who the peer-reviewers are.

The following is the editorial workflow that every manuscript submitted to the IJCCC during the course of the peer-review process.

Every IJCCC submitted manuscript is inspected by the Editor-in-Chief/Associate Editor-in-Chief. If the Editor-in-Chief/Associate Editor-in-Chief determines that the manuscript is not of sufficient quality to go through the normal review process or if the subject of the manuscript is not appropriate to the journal scope, Editor-in-Chief/Associate Editor-in-Chief *rejects the manuscript with no further processing*.

If the Editor-in-Chief/Associate Editor-in-Chief determines that the submitted manuscript is of sufficient quality and falls within the scope of the journal, he sends the manuscript to the IJCCC Executive Editor/Associate Executive Editor, who manages the peer-review process for the manuscript.

The Executive Editor/Associate Executive Editor can decide, after inspecting the submitted manuscript, that it should be rejected without further processing. Otherwise, the Executive Editor/Associate Executive Editor assigns the manuscript to the one of Associate Editors.

The Associate Editor can decide, after inspecting the submitted manuscript, that it should be rejected without further processing. Otherwise, the Associate Editor assigns the manuscript to minimum two external reviewers for peer-review. These external reviewers may or may not be from the list of potential reviewers of IJCCC database.

The reviewers submit their reports on the manuscripts along with their recommendation of one of the following actions to the Associate Editor: Publish Unaltered; *Publish after Minor Changes; Review Again after Major Changes; Reject* (Manuscript is flawed or not sufficiently novel).

When all reviewers have submitted their reports, the Associate Editor can make one of the following editorial recommendations to the Executive Editor: Publish Unaltered; Publish after Minor Changes; Review Again after Major Changes; Reject.

If the Associate Editor recommends *"Publish Unaltered"*, the Executive Editor/Associate Executive Editor is notified so he/she can inspect the manuscript and the review reports. The Executive Editor/Associate Executive Editor can either override the Associate Editor's recommendation in which case the manuscript is rejected or approve the Associate Editor's recommendation in which case the manuscript is accepted for publication.

If the Associate Editor recommends "Review Again after Minor Changes", the Executive Editor/Associate Executive Editor is notified of the recommendation so he/she can inspect the manuscript and the review reports.

If the Executive Editor/Associate Executive Editor overrides the Associate Editor's recommendation, the manuscript is rejected. If the Executive Editor approves the Associate Editor's recommendation, the authors are notified to prepare and submit a final copy of their manuscript with the required minor changes suggested by the reviewers. Only the Associate Editor, and not the external reviewers, reviews the revised manuscript after the minor changes have been made by the authors. Once the Associate Editor is satisfied with the final manuscript, the manuscript can be accepted.

If the Associate Editor recommends *"Review Again after Major Changes"*, the recommendation is communicated to the authors. The authors are expected to revise their manuscripts in accordance with the changes recommended by the reviewers and to submit their revised manuscript in a timely manner. Once the revised manuscript is submitted, the original reviewers are contacted with a request to review the revised version of the manuscript. Along with their review reports on the revised manuscript, the reviewers make a recommendation which can be "Publish Unaltered" or "Publish after Minor Changes" or "Reject". The Associate Editor can then make an editorial recommendation which can be "Publish Unaltered" or "Review Again after Minor Changes" or "Reject".

If the Associate Editor recommends rejecting the manuscript, either after the first or the second round of reviews, the rejection is immediate.

Only the Associate Editor-in-Chief can approve a manuscript for publication, where Executive Editor/Associate Executive Editor recommends manuscripts for acceptance to the Editor-in-Chief/Associate Editor-in-Chief.

Finally, recommendation of acceptance, proposed by the Associate Editor Chief, has to be approved by the Editor-in-Chief before publication.

Instructions for authors

The papers must be prepared using a LaTeX typesetting system. A template for preparing the papers is available on the journal website http://journal.univagora.ro. In the template.tex file you will find instructions that will help you prepare the source file. Please, read carefully those instructions. (We are using MiKTeX 2.4).

Any graphics or pictures must be saved in Encapsulated PostScript (.eps) format.

Papers must be submitted electronically to the following address: ccc@univagora.ro. You should send us the LaTex source file (just one file - do not use bib files) and the graphics in a separate folder. You must send us also the pdf version of your paper.

The maximum number of pages of one article is 20. The publishing of a 12 page article is free of charge (including a bio-sketch). For each supplementary page there is a fee of 50 Euro/page that must be paid after receiving the acceptance for publication. The authors do not receive a print copy of the journal/paper, but the authors receive by email a copy of published paper in pdf format.

The papers must be written in English. The first page of the paper must contain title of the paper, name of author(s), an abstract of about 300 words and 3-5 keywords. The name, affiliation (institution and department), regular mailing address and email of the author(s) should be filled in at the end of the paper. Manuscripts must be accompanied by a signed copyright transfer form. The copyright transfer form is available on the journal website.

Please note: To avoid unnecessary delays in publishing you are kindly asked to consider all recommendations expressed in the template. We do not accept submissions in other formats (pdf only, Microsoft Word, etc).

Checklist:

1. Completed copyright transfer form.
2. Source (input) files.
    - One LaTeX file for the text.
    - EPS files for figures in a separate folder.
3. Final PDF file (for reference).

Order

If you are interested in having a subscription to "Journal of Computers, Communications and Control", please fill in and send us the order form below:

ORDER FORM		
I wish to receive a subscription to "Journal of Computers, Communications and Control"		
NAME AND SURNAME:		
Company:		
Number of subscription:	Price Euro	for issues yearly (4 number/year)
ADDRESS: City: Zip code: Country: Fax: Telephone: E-mail:		
Notes for Editors (optional)		

1. Standard Subscription Rates for Romania (4 issues/2007, more than 400 pages, including domestic postal cost): 90 EURO.

2. Standard Subscription Rates for other countries (4 issues/2007, more than 400 pages, including international postal cost): 160 EURO.

For payment subscription rates please use following data:
HOLDER: Fundatia Agora, CUI: 12613360
BANK: BANK LEUMI ORADEA
BANK ADDRESS: Piata Unirii nr. 2-4, Oradea, ROMANIA
IBAN ACCOUNT for EURO: RO02DAFB1041041A4767EU01
IBAN ACCOUNT for LEI/ RON: RO45DAFB1041041A4767RO01
SWIFT CODE (eq. BIC): DAFBRO22
Mention, please, on the payment form that the fee is "for IJCCC".

EDITORIAL ADDRESS:

CCC Publications
Piata Tineretului nr. 8
ORADEA, jud. BIHOR
ROMANIA
Zip Code 410526
Tel.: +40 259 427 398
Fax: +40 259 434 925
E-mail: ccc@univagora.ro, Website: www.journal.univagora.ro

Copyright Transfer Form

To The Publisher of the International Journal of Computers, Communications & Control

This form refers to the manuscript of the paper having the title and the authors as below:

The Title of Paper (hereinafter, "Paper"):

. . . . . .

The Author(s):

. . . . . .

. . . . . .

. . . . . .

. . . . . .

The undersigned Author(s) of the above mentioned Paper here by transfer any and all copyright-rights in and to The Paper to The Publisher. The Author(s) warrants that The Paper is based on their original work and that the undersigned has the power and authority to make and execute this assignment. It is the author's responsibility to obtain written permission to quote material that has been previously published in any form. The Publisher recognizes the retained rights noted below and grants to the above authors and employers for whom the work performed royalty-free permission to reuse their materials below. Authors may reuse all or portions of the above Paper in other works, excepting the publication of the paper in the same form. Authors may reproduce or authorize others to reproduce the above Paper for the Author's personal use or for internal company use, provided that the source and The Publisher copyright notice are mentioned, that the copies are not used in any way that implies The Publisher endorsement of a product or service of an employer, and that the copies are not offered for sale as such. Authors are permitted to grant third party requests for reprinting, republishing or other types of reuse. The Authors may make limited distribution of all or portions of the above Paper prior to publication if they inform The Publisher of the nature and extent of such limited distribution prior there to. Authors retain all proprietary rights in any process, procedure, or article of manufacture described in The Paper. This agreement becomes null and void if and only if the above paper is not accepted and published by The Publisher, or is withdrawn by the author(s) before acceptance by the Publisher.

Authorized Signature (or representative, for ALL AUTHORS): . . . . . .

Signature of the Employer for whom work was done, if any: . . . . . .

Date: . . . . . .

Third Party(ies) Signature(s) (if necessary): . . . . . .