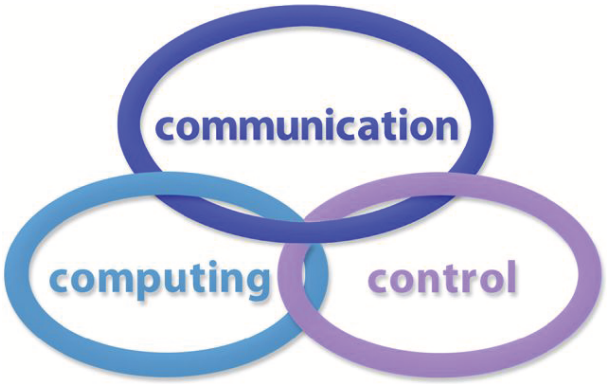
INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836



A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2015 Volume: 10 Issue: 4 (August)

This journal is a member of, and subscribes to the principles of, the Committee on Publication Ethics (COPE).

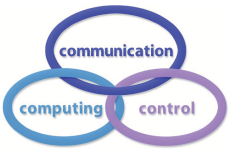
CCC Publications - Agora University Editing House

CCC Publications

http://univagora.ro/jour/index.php/ijccc/

BRIEF DESCRIPTION OF JOURNAL

Publication Name: International Journal of Computers Communications & Control.
Acronym: IJCCC; **Starting year of IJCCC:** 2006.
Abbreviated Journal Title in JCR: INT J COMPUT COMMUN.
International Standard Serial Number: ISSN 1841-9836.
Publisher: CCC Publications - Agora University of Oradea.
Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).
Founders of IJCCC: Ioan DZITAC, Florin Gheorghe FILIP and Mişu-Jan MANOLESCU.
Logo:



Indexing/Coverage:

- Since 2006, Vol. 1 (S), IJCCC is covered by Thomson Reuters and is indexed in ISI Web of Science/Knowledge: Science Citation Index Expanded.
- Journal Citation Reports (JCR - Science Edition), IF = 0.694 (JCR2013).
  Subject Category:
  - Automation & Control Systems: Q4 (46 of 59);
  - Computer Science, Information Systems: Q3 (96 of 135).
- Since 2008, 3(1), IJCCC is covered in Scopus, SJR2013 = 0.231, H index = 13.
  Subject Category:
  - Computational Theory and Mathematics: Q4;
  - Computer Networks and Communications: Q3;
  - Computer Science Applications: Q3.
- Since 2007, 2(1), IJCCC is covered in EBSCO.

Focus & Scope: International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computer and control from the universities, research units and industry.

To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of original scientific papers that focus on the integration of the 3 "C" (Computing, Communication, Control).

In particular the following topics are expected to be addressed by authors:

- Integrated solutions in computer-based control and communications;
- Computational intelligence methods (with particular emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm intelligence);
- Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

Editorial Address:
Agora University/ R&D Agora Ltd. / S.C. Cercetare Dezvoltare Agora S.R.L.
Piata Tineretului 8, Oradea, jud. Bihor, Romania, Zip Code 410526
Tel./ Fax: +40 359101032
E-mail: ijccc@univagora.ro, rd.agora@univagora.ro, ccc.journal@gmail.com
Journal website: http://univagora.ro/jour/index.php/ijccc/

Gheorghe Păun
Institute of Math. of Romanian Academy
Bucharest, PO Box 1-764, Romania
gpaun@us.es

Mario de J. Pérez Jiménez
Dept. of CS and Artificial Intelligence
University of Seville, Sevilla,
Avda. Reina Mercedes s/n, 41012, Spain
marper@us.es

Dana Petcu
Computer Science Department
Western University of Timisoara
V.Parvan 4, 300223 Timisoara, Romania
petcu@info.uvt.ro

Radu Popescu-Zeletin
Fraunhofer Institute for Open
Communication Systems
Technical University Berlin, Germany
rpz@cs.tu-berlin.de

Imre J. Rudas
Óbuda University
Budapest, Hungary
rudas@bmf.hu

Yong Shi
School of Management
Chinese Academy of Sciences
Beijing 100190, China &
University of Nebraska at Omaha
Omaha, NE 68182, U.S.A.
yshi@gucas.ac.cn, yshi@unomaha.edu

Athanasios D. Styliadis
University of Kavala
Institute of Technology
65404 Kavala, Greece
styliadis@teikav.edu.gr

Gheorghe Tecuci
Learning Agents Center
George Mason University
U.S.A.
University Drive 4440, Fairfax VA
tecuci@gmu.edu

Horia-Nicolai Teodorescu
Faculty of Electronics and
Telecommunications
Technical University "Gh. Asachi" Iasi
Iasi, Bd. Carol I 11, 700506, Romania
hteodor@etc.tuiasi.ro

Dan Tufiş
Research Institute for Artificial Intelligence
of the Romanian Academy
Bucharest, "13 Septembrie" 13, 050711,
Romania
tufis@racai.ro

Lotfi A. Zadeh
Director,
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division
University of California Berkeley,
Berkeley, CA 94720-1776
U.S.A.
zadeh@eecs.berkeley.edu

DATA FOR SUBSCRIBERS
Supplier: Cercetare Dezvoltare Agora Srl (Research & Development Agora Ltd.)
Fiscal code: 24747462
Headquarter: Oradea, Piata Tineretului Nr.8, Bihor, Romania, Zip code 410526
Bank: BANCA COMERCIALA FEROVIARA S.A. ORADEA
Bank address: P-ta Unirii Nr. 8, Oradea, Bihor, România
IBAN Account for EURO: RO50BFER248000014038EU01
SWIFT CODE (eq.BIC): BFER

INTERNATIONAL CONFERENCE on ORIENTAL THINKING and FUZZY LOGIC
(Celebration of the 50th Anniversary of Fuzzy Sets in Big Data Era)
Dalian, China, August17-20, 2015

Organizing Institutes
Fuzzy Information & Engineering Branch, Operation Research Society of China (FIEB, ORSC).
Technical Committee on Business Intelligence,Chinese Society for Management Modernization (TCBI, CSMM).
Extenics Society, Chinese Association for Artificial Intelligence (ES, CAAI).
Liaoning Technical University, China (LNTU).
Dalian University, China.
Dalian Polytechnic University, China.
Beijing Cazl Technology Service Co., Ltd.

In order to commemorate the 50th anniversary of Fuzzy Sets and Systems and cope with the challenges of Big Data Tide, we hold the International Conference on Oriental Thinking and Fuzzy Logic (Celebration of the 50th Anniversary of Fuzzy Sets in Big Data Era) during August 17-20th, 2015 in the famous and beautiful coastal city, Dalian, China. This conference is a united conference of the 8th International Conference on Fuzzy Information and Engineering held by FIEB, ORSC, the 2nd Symposium on Big Data and Data Science held by TCBI, CSMM, Forum on Extenics and Innovation Methods held by ES, CAAI and the 3rd Workshop on Intelligence Engineering and Mathematics held by LNTU.

The honorary chair for this conference will be the founder of Fuzzy Sets Theory, Prof. Lotfi A. Zadeh. His friend, Prof. Peizhuang Wang, who is the disseminator of fuzzy mathematics and its applications in China, will be the chair of the Advisory committee.

Prof. L. A. Zadeh will deliver a video speech for the conference. This conference will review the significant contributions which Fuzzy Sets Theory has made to the Data Science, and will welcome new theory and methods. Prof. Peizhuang Wang will introduce the Factor Space Theory created by Chinese scholars in his keynote speech.

Taking this opportunity, we sincerely welcome our colleagues worldwide to join us for this conference and share the latest and excited progress.

International Program Committee: *Chair*: Yong Shi;
Members: Bingyuan Cao (China), Shuili Chen (China), Yixiang Chen (China), Yongyi Chen (China), Liya Ding (China), Ioan Dzitac (Romania), Enrique Herrera Viedma (Spain), Jiali Feng (China), Qiang Fu (China), Minghu Ha (China), Liyan Han (China), Qing He (China), Baoqing Hu (China), Weidong Hu (China), Chongfu Huang(China), Gang Kou (China), Jianping Li (China), Hongxing Li (China), Taifu Li (China), Chenguang Lu(China), H. C. Lui (USA), Shengquan Ma (China), Honghai Mi (China), Duoqian Miao (China), Zhiwen Mo (China), He Ouying (China), Xiantu Peng (USA), Zhiquan Qi (China), Yan Shi (China), Zhenming Song(China), Yingjie Tian (China), Shaocheng Tong (China), LidongWang (China), Xizhao Wang(China), Xueping Wang (China), Ciyuan Xiao (China), Xiangjun Xie (China), Yang Xu (China), Fangqu Xu (China), Xiaoquan Xu (China), Zeshui Xu (China), Guojun Yan (China), Chunyan Yang (China), Liangzhong Yi (China), Ye Yin(China), Mingsheng Ying (China), Fusheng Yu (China), Xuehai Yuan (China), Wenyi Zeng (China), Bo Zhang (China), Chengyi Zhang (China), Changqing Zhang (USA), Qiang Zhang, Bin Zhao.

Honorary Chair: Lotfi A. Zadeh.

Conference Chairs: Zengliang Liu and Peizhuang Wang.

Organizing Committee Chairs: Jiren Wang and Kaiqi Zou.

7th IFAC CONFERENCE on MANAGEMENT and CONTROL of PRODUCTION and LOGISTICS (MCPL 2016)

Bremen, Germany, February 22-24, 2016

http://www.mcpl2016.logdynamics.de

The 7th IFAC Conference on Management and Control of Production and Logistics (MCPL 2016) will be held in Bremen (Germany) from 22nd to 24th of February 2016 jointly with the 5th International Conference on Dynamics in Logistics (LDIC 2016) and several satellite events. Accepted papers will be published in the proceedings of the event using the open-access IFAC-PapersOnLine.

Scope of the Conference

The conference, sponsored by IFAC, aims to bring together researchers and practitioners from different areas of production and logistics with a special focus on the engineering side of management and control of such systems. The central idea is to establish a common ground in order to promote a synergy among different disciplines for exploring new solutions for complex scientific and technical challenges. The objectives of the conference are to provide high quality research and professional interactions for the advancement of science, technology and fellowship. It also provides the participants an opportunity to present their research papers and experience reports, and to take part in open discussions.

Topics

Topics of interest include, but are not limited to: *Modeling and Simulation; Decision-Support Systems: Concepts, Methods and Algorithms; Discrete Event Systems; Cyber-physical Production and Logistic Systems; Probabilistic and Statistical Modeling; Production Planning and Scheduling; Operational Research Applications; Control Methods and Concepts; Robotics and Man-Machine Interaction; Factory Automation; Intelligent Manufacturing Systems; Advanced Process Control and Wireless Automation; Lean Six Sigma: Enterprise, Manufacturing and Healthcare; ERP and Inventory Control; Management of Organizations; Supply Chain and Green Supply Chain Management; Urban Freight Distribution and City Logistics; Information Technology in Production, Logistics and Management; Humanitarian Logistics; Socio-technical and Cognitive Aspects in Manufacturing and Logistics; Quality Management Systems and Performance Indicators.*

MCPL

The IFAC MCPL 2016 is the 7th in a very successful series of events, previously held in Fortaleza (Brazil), Campinas (Brazil), Grenoble (France), Santiago (Chile), Sibiu (Romania) and Coimbra (Portugal). This seventh edition will be organized by the BIBA Bremer Institut für Produktion und Logistik, one of the most important research centers for Production and Logistic Systems in Europe. The conference will be held in the Hanseatic City Bremen on the banks of the River Weser, one of the biggest logistics hubs in Europe. The city combines high-tech and picturesque narrow streets built in centuries past, and its rich heritage of history is greatly cherished and lovingly preserved. Bremen is the only city in Germany to have an airport ten minutes away from the city. Destinations like London, Madrid, Munich or Vienna can be reached easily by plane. Submissions The conference submission tool "PaperCept" will open in July 2015. Papers submitted to the main conference must contain original research and should not exceed six pages. Simultaneous submission to other conferences with proceedings or submission of material that has already been published elsewhere is not allowed.

Program Chairs: Jurgen Pannek (pan@biba.uni-bremen.de)
and Florin Gheorghe Filip (ffilip@acad.ro).

6th INTERNATIONAL CONFERENCE on COMPUTERS, COMMUNICATIONS and CONTROL (ICCCC 2016)

Hotel President, Baile Felix, Oradea, Romania, May 10-14, 2016

Organized by Agora University of Oradea,

under the aegis of Romanian Academy: Information Science and Technology Section.

http://univagora.ro/en/icccc2016/

Scope and Topics

The International Conference on Computers Communications and Control (ICCCC) has been founded in 2006 by I. Dzitac, F.G. Filip and M.-J. Manolescu and organized every even year by Agora University of Oradea, under the aegis of the Information Science and Technology Section of Romanian Academy and IEEE - Romania Section.

The goal of this conference is to bring together international researchers, scientists in academia and industry to present and discuss in a friendly environment 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.

In particular the following topics are expected to be addressed by authors:

1) Integrated solutions in computer-based control and communications;
2) Network Optimization and Security;
3) Computational intelligence methods (with particular emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm intelligence);
4) Data Mining and Intelligent Knowledge Management;
5) Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies);
6) Membrane Computing - Theory and Applications;
7) Stereovision Based Perception for Autonomous Mobile Systems and Advanced Driving Assistance.

Special Sessions

Special Session 1: Network Optimization and Security, Organizer and Chair: Yezid DONOSO, University de los Andes, Colombia;

Special Session 2: Data Mining and Intelligent Knowledge Management, Organizers and Chairs: Gang KOU and Yi PENG, China;

Special Session 3: Computational Intelligence Methods, Organizers and Chairs: Razvan ANDONIE and Donald DAVENDRA, Central Washington University, USA;

Special Session 4: Advanced Decision Support Systems, Organizer and Chair: Marius CIOCA, Lucian Blaga University of Sibiu, Romania;

Special Session 5: Fuzzy Control, Modeling and Optimization, Organizer and Chair: Radu-Emil PRECUP, Politehnica University of Timisoara, Romania;

Special Session 6: Membrane Computing - Theory and Aplications, Organizers and Chairs: Marian GHEORGHE (UK) and Florentin IPATE (Romania);

Special Session 7: Stereovision Based Perception for Autonomous Mobile Systems and Advanced Driving Assistance, Organizer and Chair: Sergiu NEDEVSCHI, Technical University of Cluj-Napoca, Romania.

Keynote Speakers: Enrique HERRA VIEDMA(Spain), Zenonas TURSKIS (Lithuania), Gang KOU (China).

Conference Chairs: Ioan DZITAC, Florin Gheorghe FILIP and Misu-Jan MANOLESCU.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, Volume 10, Issue 4, August, 2015.

CONTENTS

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):463-470, August, 2015.

Alleviation of Binding Update Re-registration Handoff Latency at Home Agent Failure in MIPv6 Network

A. Avelin Diana, K. Sundarakantham, S. Mercy Shalinie

A. Avelin Diana*, K. Sundarakantham, S. Mercy Shalinie
Department of Computer Science and Engineering
Thiagarajar College of Engineering, Madurai
dianaavelin@gmail.com, kskcse@tce.edu, shalinie@tce.edu
*Corresponding author: dianaavelin@gmail.com

Abstract: Home Agent (HA) is an indispensable entity for binding connectivity to route packets between Mobile Node (MN) and Correspondent Node (CN). MIPv6 allows the deployment of redundant HAs to overcome HA failure. Different approaches resolve this issue to recuperate binding association information. This paper compares the effect of handoff latency in various methods and proposes a Reliable HA delivery (RHAD) mechanism to mitigate the Binding Update (BU) registration latency in HA at the time of active HA failure. We use BGP domain in network architecture and apply IBGP protocol to transmit packets between Edge Router (ER) and HA. Both the theoretical evaluation and simulation results reveal that RHAD effectively reduces BU re-registration handoff latency and increases packet delivery ratio.
Keywords: MIPv6, home agent, binding update, IBGP, VHARP, Edge router.

1 Introduction

MIPv6 Networks are experiencing a rapid growth due to scarcity of IPv4 addresses and tremendous growth in portable communication devices, ubiquitous computing and mobile users. MIPv6 Network is a vital for evolving next generation internetworking systems. MIPv6 supports the mobility of Mobile Node (MN) to route packets transparently within the network [1]. MN ascertains its identity location with the Router Advertisement (RA) broadcasted from the router. HA is a router that maintains the mobility binding table to correlate the Home Address (HoA) with the Care-of-Address (CoA) of MN. Binding update list of MN maintains its binding information with HA and CN. The binding information lifetime is 420 seconds [2]. Similarly, CN contains Binding cache (BC) to store the MNs information and communicate with it directly by forming a tunnel via Route Optimization (RO) technique [1]. The RO is enhanced to tackle security, signaling overhead and handoff latency [3]. HA is responsible for the binding association of MN and CN. MN discovers HA with the secured Dynamic Home Agent Address Discovery mechanism (DHAAD) [4]. HA maintains the HA list which includes the link address of all the HAs within the network along with its preference value. MN selects the HA based on the higher preference value [5]. ICMP DHAAD message contains the preference value which gets decremented by one as it propagates along each router. HA list between HAs are sustained by different protocols like HARP [6], VHARP [7], HA-to-HA protocol [8]. HAs experience single point of failure and it results in bottleneck within the network [9]. A single HA failure causes loss of MNs binding association information. To overcome such problems, MIPv6 allows the deployment of multiple HAs with the redundant HA to backup the bindings at the time of failure [10]. This paper evaluates the BU registration handoff latency of various approaches during HA failure and proposes a RHAD mechanism to mitigate the registration delay of MN with the new active HA. This method reduces the BU re-registration handoff latency and improves Quality of Service (QoS) in MIPv6 network.

This paper is organized as follows: Section 2 discusses the related work in this area. Section 3 proposes the network architecture of RHAD mechanism with the theoretical evaluation. Section 4 provides the simulated results and finally we conclude this paper in Section 5.

2 Related Work

HA failure results in service connectivity breakdown. Redundant HA can avail uninterrupted service by the capture of BC contents from active HA. The standby HA detects the active HA failure with the periodic Hello messages [6]. Different solutions have been proposed to solve the HA failure issue. Secondary HAs from various home link takeover the service of primary HA and the registration delay is high in this approach [11]. In [12], the redundant HAs are in the same home link and backup the data of failed primary HA. Redundant HA of [13] and [14] follow the same approach as said in [12]and additionally it considers load balancing issues. Primary and Secondary HAs are synchronized with the transport layer connections and the registration delay is less in [13]. Efficient fault tolerant protocol provides a stable storage method in which the failure of a mobility agent is recovered by another mobility agent using check pointing and receiver based pessimistic message logging approach [15]. The mobility agent of this approach stores all the MNs binding.

In quorum based mechanism, mobility binding of MN is stored in the backup quorum of network segment [16]. Here, every HA finds a new HA as its standby HA. All the HAs of virtual home agent method share one global address and only one HA is set active [17]. Redundant HA set is a collection of active and standby HAs in which the failure recovery is maintained by HA-Virtual Switch (HA-V) and HA Hard Switch (HA-H). In HA-V the MN executes IKE exchange on standby HA with Home Address (HoA) assignment but in HA-H it exchanges without HoA. At the time of active HA failure, standby HA sends home agent switch message to all the registered MNs to redirect the BU messages [7]. In VHARP, the virtual home agent address is activated by standby HA and every state of MN is to be synchronized. The backup HAs are considered from different home links in VHAHA [18]. The home agents are arranged 2 in a chain and backup its binding to the adjacent home agent in [19].In Mobile IPv6 the message exchange is twice than that of HARP and VHARP since theMN does not involve in HA failure detection and recovery [10].

In all the above approaches, the MN should update its registration to the redundant HA at the time of active HA failure. Our RHAD technique reduces the registration delay since the MN is not essential for this registration process. The connectivity between MN to a router in the network is wireless, so the packet delivery ratio is also improved by RHAD.

3 RHAD Mechanism

3.1 RHAD Network Architecture

Our network entities include MN, HA, access router, access point, CN and Edge Router (ER) as in Figure 1. MN selects the nearest HA to reduce the probability of HA handoff failure. We create BGP domain and use IBGP to transfer the packets between MN and HA. It is noticed that the packets are never propagated to exterior network. ER is a boundary router that scrutinizes the BU message from MN and records it in router list. When the BU packet is routed to HA, first transmission occurs via ER that is nearer to MN. The stored BU information contains HoA, CoA and lifetime. It remains in the router list until the BU lifetime becomes zero. The BU is transmitted to active HA through edge router and access router. The active HA selects a standby HA from redundant HA set based on the highest preference value and back up its data. The

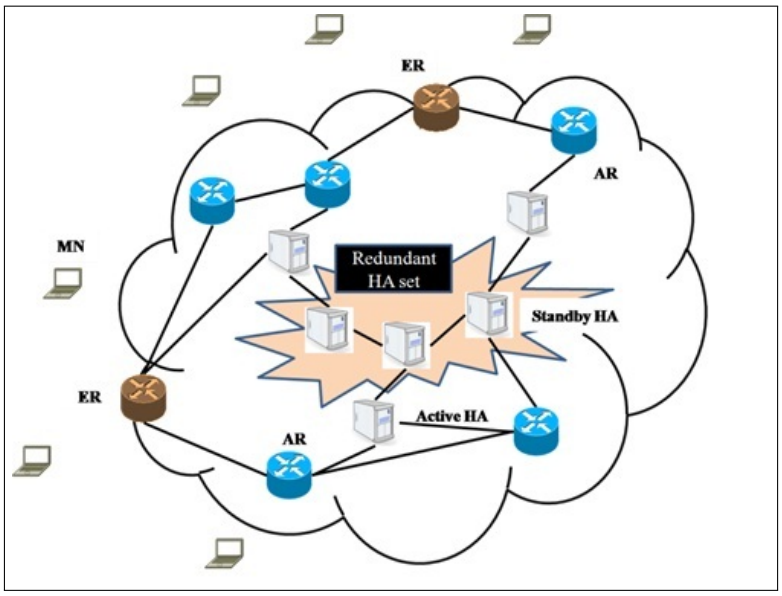


Figure 1: RHAD Network

redundant HA is provided with virtual HoA and it should be noticed that the binding should be synchronized between active and standby HA. In our approach we combine VHARP and BGP protocol to mitigate the failure recovery handoff latency. VHARP is transparent to MN and it maintains the HA list within the network. At the time of active HA failure, the edge router transmits BU information to the redundant HA.

3.2 Theoretical Evaluation

The BU re-registration handoff ow using RHAD approach is illustrated in Figure 2. The total message ow handoff time includes processing time and transmission time. Message retransmission is unnecessary in wired medium. Wireless link is least stable since message loss can occur at any instance and therefore message retransmission becomes indispensable. The transmission of BU message between MN and edge router is wireless. The additional signal processing time is evaluated below.

$$T_{BU}^{1} = \sum_{i=n}^{\infty} T_{BU}\,(n_f) \cdot prob(n_f\ failure\ and\ one\ success) \tag{1}$$

n_f represents the number of link failures. If the BA is not received after sending BU request we assume that message is lost and hence retransmitted. If n_f failure occurs then $T_{out}$and message retransmission takes n_f times. Usually, T_{out} is observed as 2 ms [13].

$$T_{BU}(n_f) = P_w + n_f \cdot (T_{out} + P_{wl}) \tag{2}$$

From (1) and (2),

$$T_{BU}^{1} = \Sigma_n^{\infty}\,\{P_{wl} + n_f \cdot (T_{out} + P_{wl})\} \cdot prob(n_f\ failure\ and\ one\ success) \tag{3}$$

$\Sigma_n^{\infty} n_f \cdot prob\,(n_f\ failure\ and\ one\ success)$ is obtained from infinite geometric progression.

$$\Sigma_n^{\infty} n_f \cdot prob\,(n_f\ failure\ and\ one\ success) = \frac{r}{r-1} \tag{4}$$

Here, r represents the link probability failure and it has the value 0.5 [13].

$$T_{BU}^{1} = P_{wl} + (T_{out} + P_{wl}) \frac{0.5}{1 - 0.5} \tag{5}$$

$$T_{BU}^{1} = 2P_{wl} + T_{out} \tag{6}$$

Similarly, the BA received from edge router to MN is obtained as

$$T_{BA}^{2} = 2P_{wl} + T_{out} \tag{7}$$

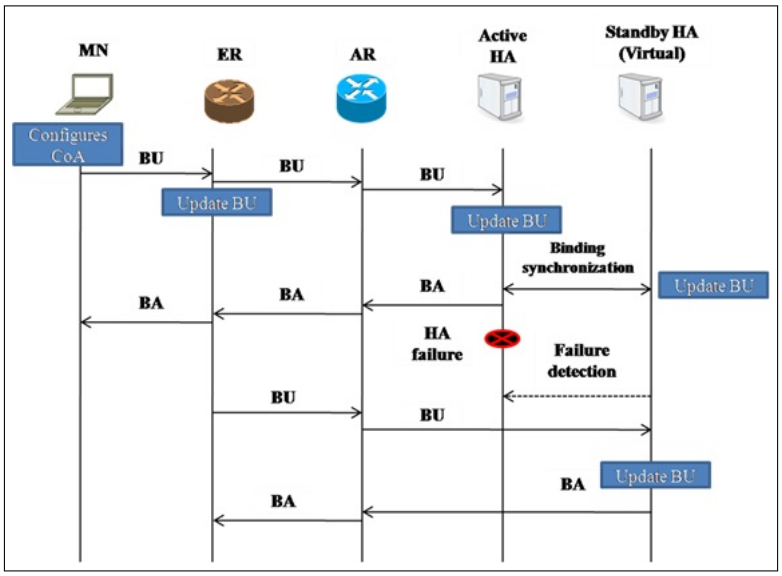


Figure 2: BU recovery registration handoff

Therefore total time for the message transmission in wireless medium is

$$T_{wl} = T_{BU}^{1} + T_{BA}^{2} \tag{8}$$

$$T_{wl} = 2\left(2P_{wl} + T_{out}\right) \tag{9}$$

The message transmission in wired medium includes the packet propagation between edge router and standby HA. The binding synchronization between active and standby home agent is achieved through VHARP protocol. The total message transmission time in wired medium is

$$T_{wi} = T_{ER-AR} + T_{AR-AHA} + T_{Bsync} + T_{AR-ER} + T_{AHA-AR} \tag{10}$$

$$T_{wi} = 2T_{ER-AR} + 2T_{AR-AHA} + T_{Bsync} \tag{11}$$

At the time of active HA failure, the total transmission time in HARP is obtained as

$$T_{HARP} = 2T_{wl} + T_{wi} + T_{ER-AR} + T_{AR-SHA} + T_{AR-ER} + T_{SHA-AR} \tag{12}$$

$$T_{HARP} = 2\left(T_{wl} + T_{ER-AR} + T_{AR-SHA}\right) + T_{wi} \tag{13}$$

Edge router contains the information of MNs BU and hence in RHAD method the standby virtual HA acquires the BU directly from edge router and sends BA after successful delivery. Therefore the transmission between edge router and MN is not necessary.

$$T_{RHAD} = 2\,(T_{ER-AR} + \ T_{AR-SHA}) \ + \ T_{wi} + T_{wl} \tag{14}$$

We have the fixed processing time to configure CoA and to update BU message at edge router, active HA and standby HA. The processing time is same for both HARP and RHAD method.

$$PT_{RHAD} = \ PT_{CoA} + \ 4 \ PT_{BU} = 5 \ PT \tag{15}$$

The total handoff failure is obtained by summing up the transmission and processing time.

$$T^{HO}_{HARP} = T_{HARP} + PT_{HARP} \tag{16}$$

$$T^{HO}_{HARP} = 2\,(T_{wl} + T_{ER-AR} + T_{AR-SHA}) + T_{wi} + 5PT \tag{17}$$

$$T^{HO}_{HARP} = T_{RHAD} + PT_{RHAD} \tag{18}$$

$$T^{HO}_{RHAD} = 2\,(T_{ER-AR} + T_{AR-SHA}) + T_{wi} + T_{wl} + 5PT \tag{19}$$

From (7) and (8), it is clear that the BU re-registration handoff in RHAD is fast and reliable.

4 Simulation results

The system parameter for the simulation of RHAD network is based on J. Mc Nair et al. [20] as depicted in Table 1. In this experiment we analyse the BU re-registration latency of different approach. Figure 3 depicts the number of recovery messages exchanged during BU registration at the new active HA. The number of recovery messages is high in MIPv6 and other methods when compared to our RHAD approach.

Table 1: System Parameter

System parameter	**Value**
Wired link message propagation time	0.5 ms
Wireless link message propagation time	2 ms
Link failure probability	0.5
Bit rate of wired link	155 Mbps
Bit rate of wireless link	144 Kbps
Message processing time	0.5 ms

After the recovery of new HA, the MN sends BU to confirm its binding registration with it. The BU re-registration latency over time is described in Figure 5. Since this process is not initiated by MN in RHAD, the registration latency is alleviated within 35 ms for the simulation time of about 300 sec as shown in figure. The fraction of packets that suffers due to different delays in various approaches at the time of re-registration is shown in figure 6.

Packet delivery ratio is evaluated as the number of packets received by the new active HA to the total number of packets sent by the mobile node. The packet delivery ratio of diverse

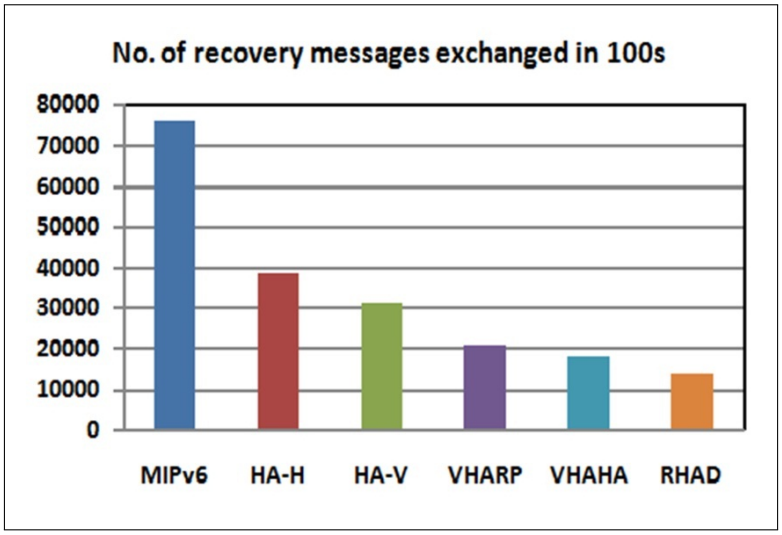


Figure 3: No of recovery messages of different approach

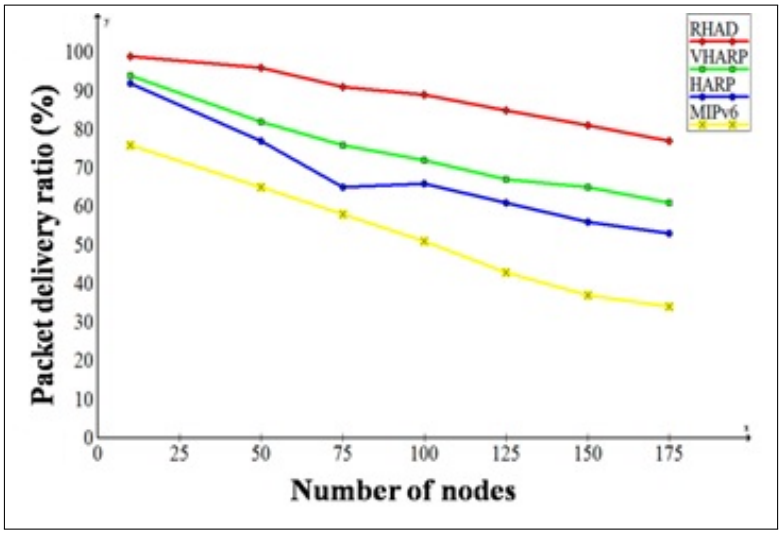


Figure 4: Active HA packet delivery ratio

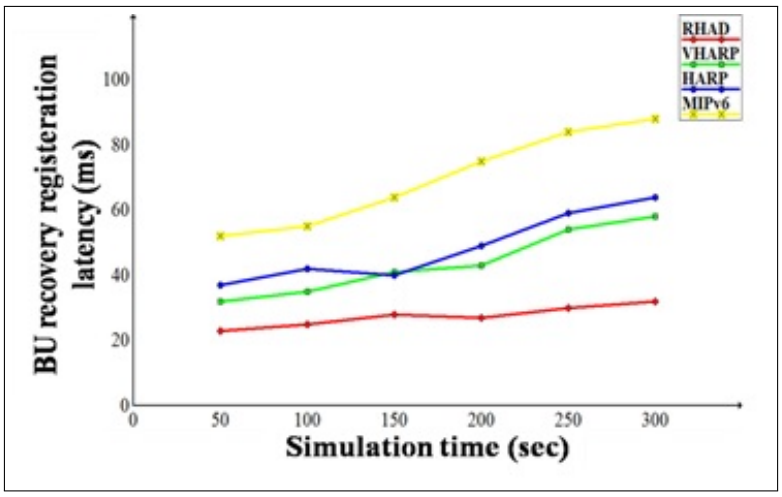


Figure 5: BU recovery registration latency

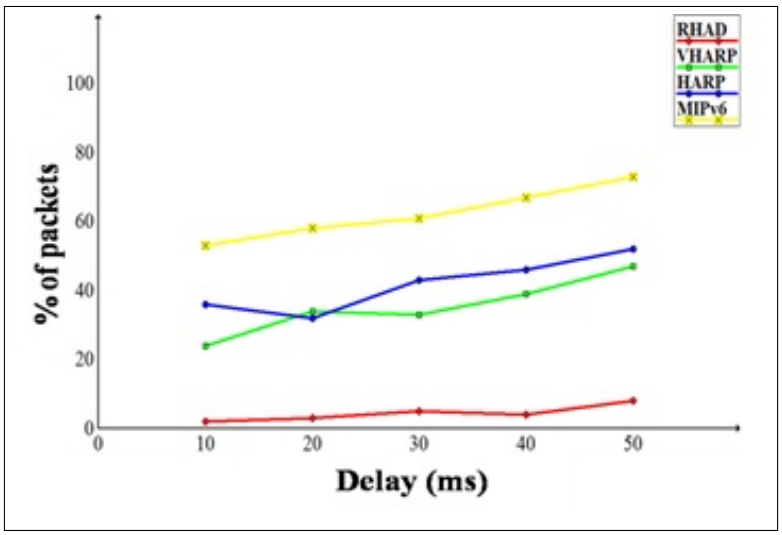


Figure 6: Suffered packets at delay

approaches is illustrated in figure 4. For 175 nodes, the percentage of packet delivery ratio in RHAD is 81% whereas for VHARP, HARP and MIPv6, it is 61%, 53% and 34% respectively. The successful packet delivery is comparatively large in our technique.

5 Conclusion

RHAD network is simulated and its handoff latency is examined. Comparatively this method mitigates the registration latency of BU after the HA failure. Our approach outperforms all the other approaches with the enlarged packet delivery ratio at HA. Currently we are analysing the HA fault tolerant mechanism and failure recovery. In future work we would like to include an efficient HA fault tolerant method to recuperate the registered BU without any loss of service connectivity.

Bibliography

[1] Johnson, D.; Perkins, C.; Arkko, T. (2004); Mobility support in IPv6, *IETF RFC 3775*, https://www.ietf.org/rfc/rfc3775.txt.

[2] http://www.kame.net

[3] Arkko, J.; Vogt, C.; Haddad, W.(2007); Enhanced Route Optimization for Mobile IPv6, Network Working Group, *RFC 4866*, https://www.rfc-editor.org/rfc/rfc4866.txt.

[4] Qian Sun et al (2004); Security Issues in Dynamic Home Agent Address Discovery, *draft-sun-mipv6-dhaadsecurity-00.txt*.

[5] http://www.cisco.com/en/US/docs/ios/ipv6/command/reference/ipv6_07.html

[6] Wakikawa, R. (2009); Home Agent Reliability Protocol, MEXT Working Group, *draft-ietf-mip6-hareliability-05.txt*.

[7] Wakikawa, R.(2011); Home Agent Reliability Protocol, MEXT Working Group, *draft-ietf-mip6-hareliability-09.txt*.

[8] Devarapalli, V.; Wakikawa, R.; Thubert, P. (2006); Local HA to HA protocol, *draft-devearapalli-mip6-nemo-local-haha-01.txt.*

[9] Ng, C.; Thubert, P.; Watari, M.; Zhao, F. (2007); Network Mobility Route Opti-mization Problem Statement, Network Working Group, *RFC 4888.*

[10] SeungSun Hong (2004); Multihoming Scenarios with Multiple Home Agents in Mobile IPv6, *draft-hong-mip6-multihoming-scenarios-00.txt.*

[11] Wakikawa, R.; Devarapalli, V.; Thubert, P. (2003); Inter Home Agents Protocol (HAHA), IETF Draft, *draft-wakikawamip6- nemo-haha-00.txt.*

[12] Deng (2003); Load balance for Distributed HAs in Mobile IPv6, IETF Draft, *draft-wakikawamip6-nemo-haha-00.txt.*

[13] Heissenhuber, F.; Fritsche, W.; Riedl, A. (1999); Home Agent Redundancy and Load Balancing in Mobile IPv6, *5th International Conference Broadband Communications*, 235-244.

[14] Faizan, J.; El-Rewini, H.; Khalil, M. (2005); VHARP: Virtual Home Agent Reli-ability Protocol for Mobile IPv6 based Networks, *Int. conf. on Wireless Networks, Communications, and Mobile Computing*, DOI:10.1109/WIRLES.2005.1549599, 1295-1300.

[15] Ahn, J.H.; Hwang, C.S. (2011); Efficient Fault-Tolerant Protocol for Mobility Agentsin Mobile IP, *International Parallel and Distributed Processing Symposium*, DOI:10.1109/IPDPS.2001.925103, 1273-1280.

[16] Chen, Y.S.; Chen, C.H.; Fang, H.Y. (2008); An Efficient Quorum-Based Fault-Tolerant Approach for Mobility Agents in Wireless Mobile Networks, *IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC)*, 373-378.

[17] Faizan, J.; Rewini, H.E. et al (2006); Introducing Reliability and Load Balancing in Mobile IPv6 Based Networks, *Journal of Wireless Communications and Mobile Computing*, 6:1-19.

[18] Rathi, S.; Thanushkodi, K. (2006); Design and Performance Evaluation of anEffcient Home Agent Reliability Protocol, *International Journal of Recent Trends in Engineering*, 2(1):26-32.

[19] Yujun Zhang; Hanwen Zhang (2011); A Mobile Agent Fault-Tolerant Method Based on the Ring Detection Backup Chain for Mobile IPv6 Networks, *IEEE Int. Conf. on Communications*, DOI: 10.1109/icc.2011.5962677, 1-5.

[20] McNair, J.; Akyildiz, I.F.; Bender, M.D. (2001); Handoffs for real-time traffc in mobile IP version 6 networks, *IEEE GLOBECOM*, DOI:10.1109/GLOCOM.2001.966325, 6: 3463-3467.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):471-479, August, 2015.

Efficient Design and Deployment of Aqua Monitoring Systems Using WSNs and Correlation Analysis

S. Babu Chandanapalli, E. Sreenivasa Reddy, D. Rajya Lakshmi

Suresh Babu Chandanapalli*
Associate Professor, Department of CSE,
Gudlavalleru Engineering College, Gudlavalleru-521356,
Andhrapradesh, India.
*Corresponding author: sureshdani2004@gmail.com

Dr E. Sreenivasa Reddy
Professor of CSE, ANU College of Engineering & Technology,
Acharya Nagarjuna University, Guntur-522510, Andhrapradesh, India.
edara_67yahoo.com

Dr D. Rajya Lakshmi
Professor, HOD of CSE, JNTUK University College of Engineering,
Vizianagaram- 535002, Andhrapradesh, India.
jrajyalashmi@gmail.com

Abstract: The roots of innovation are extending towards every field to provide ace solution. We cater an ace solution for aquaculture, where their yields (shrimp, fish, etc.) depends on the ponds water characteristics. The parameters depending on water must kept at certain optimal levels for better cultivation of Aqua. The parameters of water extremely project alterations during the day and also alter depending upon the environmental conditions i.e., it is necessary to monitor these parameters with high frequency. We adopt wireless sensor networks to monitor aqua forms. This system consists of two modules, they are transmitter and receiver station. We navigate data to database at receiver station through the GSM. The graphical user interface was designed in such a manner that the observations are forwarded to the farmer as message in their respective local languages to their mobile phones. That alerts them in unhygienic environmental conditions for adopting suitable measures.
Keywords: Aquaculture, wireless sensor networks, IAR-Kick, pH;

1 Introduction

Aquaculture is one of the widely extending industry attributable to the rapid demand for fish and seafood all over the world. The term aquaculture is referred as the cultivation of fish,plants and animals in various types of environments that includes rivers, ponds and oceans. Aquaculture consists of two types i.e. one is marine aquaculture that is nothing but the cultivation of species in ocean and another is freshwater aqua culture where species are cultivated in native water bodies. In shrimp culture, it is observed that samples that are taken into consideration to predict low levels in dissolved oxygen, temperature, salinity and PH levels. Taking all these parameters into consideration deploying sensors in shrimp culture for monitoring water quality and alert regarding contaminants in water will yield exceptional results. [1] [8]. The analysis of water quality desires consistent observation of water depended parameters in significant catchments. The various parameters which we consider as pH, dissolved oxygen, water temperature and water levels [2] at various depths. Making use of various ideal instruments like sensors and wireless sensor networks will produce better results. The adoption of Zigbee standard for short range and low cost module of wireless sensor network is developed in real time information system [3], in

which it consists of small sensor nodes, coordinator or gateway node and personal computer. In this system the smart sensor node monitor the ecological parameters such as water level, pH and dissolved oxygen and transmits it to the coordinator or gateway node from which data is again transferred to the personal computer where it is visualized for human-computer interface. The application of wireless sensor networks in various fields for detection of ecological parameters and transferring data to database using network. Yet, the wireless sensor networks have many constraints like memory, processing power and limited batter energy but the efficient utilization of energy is a decisive issue. We propose the implementation of wireless sensor networks to have a distributed collection of sensor nodes networked together to transfer the raw data to a central location known as base station through GSM. Every sensor node consists of a micro-controller, some sensors and a radio transceiver for communication [3]. The micro-controller is used for in-network processing for transferring needed information instead of raw data. The information which is transferred is preserved in a database and analyzed for further process. After analysis the data from database is forwarded to the farmers as a message to their mobile in their respective languages to alert them about the unhygienic environmental conditions. The proposed system makes farmers aware about the vulnerabilities so as to resolve them.

2 Motivation

Abrupt change in climate proves fatal for fish

G. Nagaraju

ELURU: The extremities in weather conditions delivered a deadly blow on Jampana Rama Raju, a fish farmer from Tedinadu village in Undi mandal of West Godavari district. Around 5-6 tonnes of the fish in his 100-acre pond worth Rs. 5 lakh died in the last two days due to deficiency in oxygen levels caused by sudden changes in the climate.

"I have just buried the dead fish in my fields. I could not have recovered even the fuel expenses had I brought them to the market at Akiveedu," he said.

According to Mr. Rama Raju, over 400 tonnes of fish died in his village alone in 2,000 acres of fishponds.

"It all happened when all the fish got ready for harvest and I was planning to catch them within a week," laments Mr Satyanarayana Raju who raised fish in a five-acre pond by investing Rs. 1.25 lakh per acre. He dumped the dead fish on the roadside.

The dead fish was sold at a price as low as Rs 2-5 per kg on Monday. The Akiveedu market receives 15-20 tons of fish from Krishna and West Godavari districts daily. The mortality was reported mainly in the Katla and Rohu species.

Heavy loss of crop was reported in Bhimadole, Akiveedu, Nidamarru and Bhimavaram areas.

V. Ramachandra Raju, president of the Delta Fish Farmers Association, said the reported inability of the district administration to provide canal water for agriculture in the rabi also played its part in ruining the fish farmers. The aqua farmers were forced to fill their ponds with drain water let out from the agricultural fields.

The wastewater contains parasitic loads of flukes and the other bacteria which take a heavy toll on the immunity levels of the fish when a drop in the DO (deficiency of oxygen) levels is reported, according to S.K. Akbar Ali, who runs an aqua lab at Akiveedu.

COLOSSAL LOSS: A worker displays the fish which died due to drop in oxygen levels in a pond at Akiveedu in West Godavari district. - PHOTO: A.V.G. PRASAD

Figure 1: Paper cutting for problems in aquaculture

In aquaculture, the yields(shrimp, fish etc) depends on the water characteristics of the aquaculture pond. Parameters must be maintained at certain levels for better cultivation of fish yields. The parameters asPH level, Salinity, turbidity, alkalinity and nutrition level alter from day and environmental issues. From early years, considerable amount of research and experimentation have processed in this field but till there is no perfect solution of this issue. There are certain possible wireless sensor network solutions that act as a perfect solution for the problem but to their drawbacks in power utilization, configuration, communication failures, environmental influences and scalability inspired us to perform further explorations in aquaculture. [1]. Fig.2.shown given below, Certain innovations those were carried out in aquaculture for their detection in 2011 the innovation of traceability enciphers for recirculation aquaculture and in 2012 creation of water checking framework where we adopt zigbee and GPRS for transmission of gathered data [4]. In this existing system we use a PH meter where we use a probe to testing the parameters and transferring the results using wireless sensor networks. In this manual testing there are some

drawbacks as arrangement of Ph meter at some frequency,maintains of probe, time for calculation of PH and manually testing is not possible in unhygienic environmental conditions. The earlier systems are not made up of GSM for aqua monitoring system for remote connectivity and sensing parameters.

Figure 2: Manual testing of PH

2.1 Proposed System Description

We propose the automatic system for checking water characteristics of aqua ponds in hygienic and unhygienic conditions. The proposed system consists of two modules, they are transmitter and receiver station and a personal computer which is used as analyzing station. The following list contributes overall framework of the aqua monitoring system:

- Embedded system design
- Radio frequency integrated circuits.
- Wireless sensor network.
- The collected data will be analyzed by using data mining tool.
- Convert the output into local language using local language.
- User interface design in local language

2.2 Design of transmitter station

We propose the implementation of wireless sensor networks to have a distributed collection of sensor nodes networked together to transfer the raw data to a central location known as base station through GSM. Every sensor node consists of a micro-controller, sensors and radio transceiver for communication. The micro-controller is used for in-network processing for transferring needed information instead of raw data. As the sensor node are powered by battery and thereby reduction of total n umber of bits will save the battery life. Research activities are

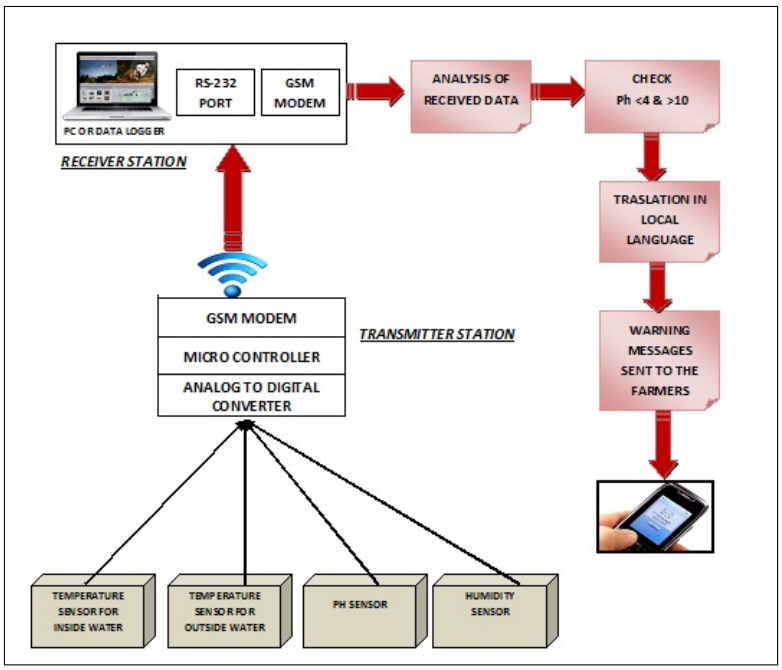


Figure 3: Overall architecture of aqua monitoring system

processing for the change of sensor node replacement for optimal coverage, topology formation, routing, data processing techniques to reduce communication costs, operating system design etc. it is also to estimate the impact of the present sensors in aquaculture as it needs few more crop cycles for generation of accurate results [6] [11]. In the proposed system, we introduce a sensor that makes use of the off-the-shelf available sensor nodes and we didnt make any specific efforts in hardware design cost reduction. The sensor is embedded in the transmitter station for monitoring of aqua forms depending on parameters like PH levels, dissolved oxygen levels, temperature inside and outside water, ammonia levels etc.The user interface was designed in VB and .NET. So that farmers and investigators may analyze and investigate the data. [7] The transmitter consists of: 1.Sensor nodes 2.GSM modems 3.Micro-controllers 4.ARM processor

2.3 Sensor nodes

Sensors are the electronic devices that sense the alterations of physical environmental conditions as sound, temperature, pressure etc.the sensor works at particular voltage and continuously transmit the signal in analog form. These signals are converted into digital form using converters. The sensors are of small size, consume low power and operate in high densities. We make use of following sensors:

- Temperature sensor for measuring the altering conditions of water
- PH sensor
- Humidity sensor.

Sensor selection

- Temperature sensor for outside water: we make use of DS18B20 I-wired digital sensor from maxim IC.

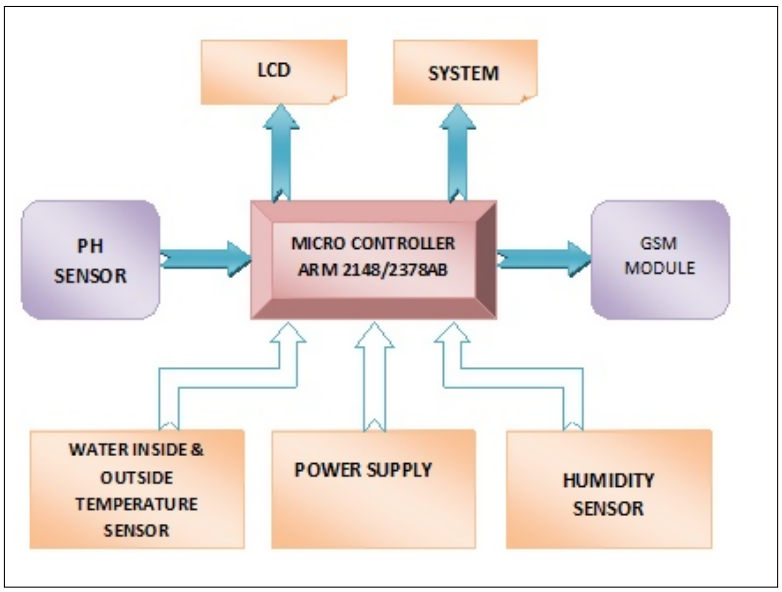


Figure 4: Block diagram of transmitter

- Temperature sensor for inside water: we make use of HOBO U-series sensor with UX 120-006M external channel data logger.
- pH sensor : we make use of a glass electrode for pH sensing.
- Humidity sensor : we use to measure the humidity.

GSM modems We make use of GSM modems for long range communication as the aquaculture ponds are far away to villages and the supply of power is not available to ponds therefore data collection nodes are not placed near to pond. We place the receiver node at data collection center in villages or towns and transmits message to the mobile phones with the help of a SIM, that supports 2G and g technologies, HSUPA, UMTS, WCDMA, HSDDA, GPRA and EDGE. These modems are embedded both at transmitter and receiver station.In this prototype we adopt AT89C52 micro-controller that is having the capability of showing high-performance by making use of CMOS-8 bit microcomputer having 8k bytes flash programmable and erasable read only memory.We make use of the ace ARM processor that comes with all the well known features and functionality.System software of the aqua monitoring system gains the sensor data and verifies with the data records and preserves the data in the database to construct a complete database. The constructed database consists of all the analyzed patterns depending on the altering environmental changes of fish ponds [5]. The sensor data is correlated with the ideal conditions present in the database. Any alteration in the ideal conditions will transmit a warning message to the farmer for adopting suitable measures. Communication process module is connected to the PC via serial port RS-232. We use IAR or RAM for controlling the wireless sensor system in the embedded software that is programmed in C. We port the hex document to the flash memory of AT-mega8 micro-controller via USB interface. The project is created with subroutine modules for sensor indicator digitization, engine transfer and value hand-off control, remote information procurement and transmission, universal and asynchronous receive transmit. The analog and digital converter of micro-controller performs cyclic scans for the sensors and considers the varying check rate that is programmable [9]. The sensor information might be contrasted and set-point qualities put away in memory. The transmission of information through the GSM to the base collector is done at serial correspondence at 9600 baud rate.

Tab.1. Technical specifications of communication protocol

Parameters	ZigBee	Bluetooth	Wi-Fi	GSM
Range	30m - 1.6 KM	30 - 3000ft	100 - 150 ft	30 - 35 KM
Power Consumption	0.2 mA(one node)	1 Watts	10.80 Watts	230 volts
Frequency	2.4G	2.4G	2.4G	900 and 1800 MHz
Cost	$350	$325	$500	$250

3 System implementation

The system was implemented in Gudlavalleru, Krishna district. The intensive aquaculture monitoring system is acquiring data from March-2014 to August-2014.this system analyzes all the appropriate operations and assesses the feasibility. We had taken a pond area of three acres and divided into four fish ponds. We deployed four sensor nodes and a transmitter node consisting of PH, Humidity and water temperature at the ponds. The whole system was tested appropriately with the connection of sensors to their modules and monitoring computer to its receiver nodes in the control room with the proper installation of sensor nodes in the fish ponds and their operations.

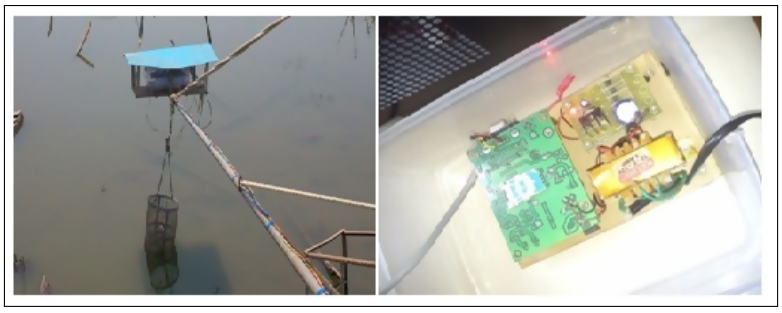


Figure 5: Installation of Transmitter station & Receiver Station at pond

4 Results and discussion

Tab.2. Summary of sensor data between Mar-2014 to July-2014

Parameters	Mar-14			Apr-14			May-14		
	Min	Max	Stdiv	Min	Max	Stdiv	Min	Max	Stdiv
*W. I.Temp.	23	26	2.12	23	25	1.4	18	22	2.83
**W .O.Temp.	25	27	1.41	25	25	0	25	26	0.7
Humidity	44.1	50.2	4.3	43.5	46	1.73	43.6	46.6	2.12
pH	6.1	6.3	0.14	5.8	6.2	0.28	5.8	6.2	0.28
Fish Mortality		1.4			1.1			1	

Parameters	June -14			July-14			Aug-14		
	Min	Max	Stdiv	Min	Max	Stdiv	Min	Max	Stdiv
*W. I.Temp.	22	24	1.41	23	23	0	23	24	0.71
**W .O.Temp.	25	27	1.41	25	26	0.71	25	27	1.41
Humidity	45.9	46.9	0.71	46.8	48.7	1.34	42	47.6	3.96
pH	5.8	6.3	0.35	5.7	6	0.21	5.7	7.8	1.48
Fish Mortality		1.2			1			1.3	

*-Water Inside Temperature **-Water Outside Temperature

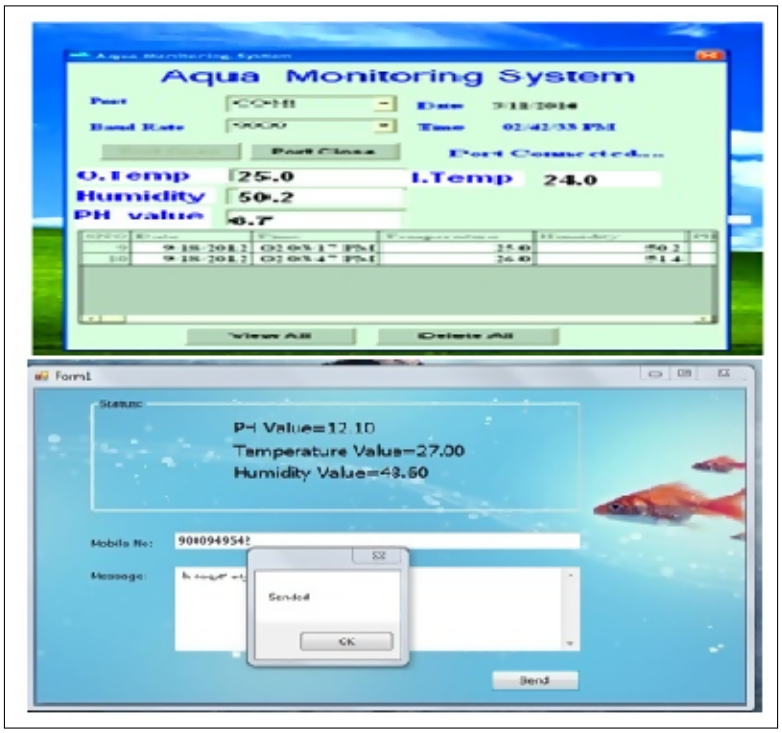


Figure 6: Monitoring screen and user interface of Aqua monitoring System

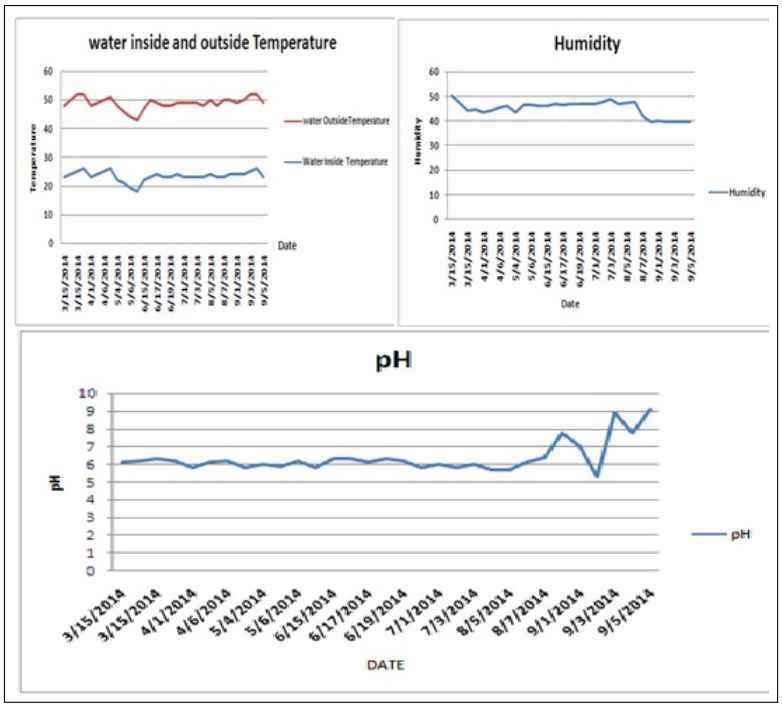


Figure 7: Water inside and outside temperature,Humidity,pH variation from a period of six months

Correlation Co-efficients

Correlation co-efficient may termed as the measure of the strength of linear association among two variables or parameters. The value of correlation always lies in between -1.0 to +1.0.if the value of correlation, C(r), is positive then we retrieve positive relationship, if negative we retrieve negative relationship [7] [10].

Formula:

$$C(r) = [N\sum XY - (\sum X)(\sum Y)]/Sqrt([N\sum X^2 - (\sum X)^2][N\sum Y^2 - (Y)^2])$$

Where,

N = The total number of values or elements
X = First Score
Y = Second Score
$\sum XY = Sumoftheproductoffirstand SecondScores$
$\sum X = SumofFirstScores$
$\sum Y = SumofSecondScores$
$\sum X^2 = SumofsquareFirstScores$
$\sum Y^2 = SumofsquareSecondScores$

Tab.3. Summary of correlation coefficients

Parameters	pH	Humidity	*W.I.Temp	**W.O.Temp
pH	1	0.75	0.85	0.95
Humidity	0.75	1	0.65	0.75
*W.I.Temp	0.85	0.65	1	0.70
**W.O.Temp	0.95	0.75	0.70	1

*-Water Inside Temperature **-Water Outside Temperature

Taking the available parameters in the table 3.i.e. pH, Water inside temperature, water outside temperature and humidity calculate the value for the variables X and Y.

5 Conclusions and Future Works

A wireless sensor network for aquaculture monitoring and control based on virtual instruments is presented. We implemented the hardware design of smart sensor nodes, transmitter/receiver with software design and prototype system of four sensor nodes with the embedding of transmitter/receiver with proper design structure. The collected data provides an accurate analysis of successful operation of the system. The proposed work can be utilized in various fields as greenhouse monitoring and control, hydrological water conservation and farm land irrigation. Future development efforts should involve enhancing the WSN by adding GPRS modules on the smart sensor nodes to monitor aquaculture environment on-line through the Internet. The monitoring program should be linked to the web server through a passive IP address in one port of the receiving computer, thus making the architecture more scalable and robust.

6 Acknowledgment

This work was supported by University Grants Commission - Minor Research Project for Teachers, Southern Eastern Regional Office, INDIA and is also sanctioned fund in-house R & D by Gudlavalleru Engineering College.

Bibliography

[1] http://www.nmfs.noaa.gov/

[2] A. Harun, D. L. Ndzi, M. F. Ramli, A. Y. M. Shakaff, M. N. Ahmad, L. M. Kamarudin, A. Zakaria, and Y. Yang (2012); Signal propagation in aquaculture environment for wireless sensor network applications, *Progress In Electromagnetics Research*, ISSN:1070-4698, E-ISSN:1559-8985, 131(1): 477-494.

[3] Zhuiykov S., Ertud K., Mirza I. (Eds.)(2010); Water Quality: Physical, Chemical and Biological Characteristics, *Sensors and Actuators B: Chemical*,ISSN: 0925-4005, Chapter 3 : 71-96.

[4] Wang, Xin; Ma, Longquan; Yang, Huizhong (2011); Online water monitoring system based on ZigBee and GPRS, *Proc. Eng.*, 15: 2680-2684.

[5] Tai, Haijiang; Liu, Shuangyin; Li, Daoliang; Ding, Qisheng; Ma, Daokun (2012); A multi-environmental factor monitoring system for aquiculture based on wireless sensor networks, *Sensor Lett.*, ISSN:1546-198X , 10 (12):265-270.

[6] D M. Razzemi, Z. Abdul Halim, M. Kusairay Musa and K. Hasbullah (2011); Development of Water Quality Monitoring System Prototype for Fresh Water Fish Culture, *Informatics Engineering and Information SciencE Communications in Computer and Information Science*, ISSN:978-3-642-25483-3 , 254(6): 470-478.

[7] Qi, J. Zhang, M. Xu, Z. Fu, W. Chen, X. Zhang (2010); *Developing WSN-based traceability system for recirculation aquaculture, Mathematical and Computer Modelling* PII: S0895-7177(10)00423-1, 53: 2162-2172.

[8] Bosma R. H., Verdegem M. C.J. (2011); Sustainable aquaculture in ponds: principles, practices and limits, *Livestock Sci.*, 139(1):58-68.

[9] Jose Juan Carbajal Hernandez, Luis P. Sanchez Fernandez, Oleksiy Pogrebnyak (2011); Assessment and prediction of the water quality in shrimp culture using signal processing techniques, *Aquaculture International*, ISSN:1573-143X, 19 (6): 1083-1104.

[10] Anita Bhatnagar, Pooja Devi(2013); Water quality guidelines for the management of pond fish culture, *International Journal of Environmental Sciences*, ISSN:0976-4402, 3(6):1981-2009.

[11] Glasgow H.B et al.(2004); Real-time remote monitoring of water quality: are view of current applications, and advancements in sensor, telemetry and computing technologies, *Journal of Experimental Marine Biology and Ecology*, ISSN: 0022-0981, DOI: 10.1016/j.jembe.2004.02.022, 300:409-448.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):480-491, August, 2015.

Development of a Fuzzy Logic System to Identify the Risk of Projects Financed from Structural Funds

M.I. Boloş, D.C. Sabău-Popa, P. Filip, A. Manolescu

Marcel Ioan Boloş*, Diana-Claudia Sabău - Popa
Departament of Finance-Accounting, Faculty of Economic Sciences,
University of Oradea
Romania, 410087 Oradea, Universitatii St. 1
marcel_bolos@yahoo.com, dpopa@uoradea.ro
*Corresponding author: marcel_bolos@yahoo.com

Petru Filip
1. Dimitrie Cantemir Christian University,
Romania, 040042 Bucharest, Splaiul Unirii, 176
2. Agora University of Oradea,
Romania, 410526 Oradea, Piata Tineretului, 8
3. University of Oradea
Romania, 410610 Oradea, University Street, 1
pfilip@uoradea.ro

Adriana Manolescu
Departament of Social Sciences
Agora University of Oradea
Romania, 410526 Oradea, Piata Tineretului, 8
adrianamanolescu@univagora.ro

Abstract: The fuzzy logic system developed in this research paper seeks to identify the financial risk of projects financed from structural funds when changes occur in project values, in the duration of the projects and in the implementation durations. Those two factors are known to influence the financial risk. The fuzzy system was simulated using Matlab and the results showed its operation and the conclusion that the financial risk of the project is dependent on the developments values and on the implementation duration. The developed and tested fuzzy logic system provides information on financial risk intensity organized into three categories: small, medium and large and on the inflection point of transition from low risk to high risk. This is considered an early warning system for the management staff with responsibilities in structural funds.
Keywords: Fuzzy Logic System (FLS), artificial intelligence, financial risk, structural funds, centroid method.

1 Introduction

The fuzzy logic systems (FLS) are used as a tool for decisions making, for the projects financed from structural funds, for the early identification of risks that affect the performance of the allocation of funds for EU member countries, through various financial instruments known as operational programs. The risk of the projects financed from structural funds has various forms, but the most important remains the financial aspect, that generates losses for the budget of the member states. Although there are now used statistical methods for risk measurement, the most common being the standard deviation (σ^2), it should be noted that they have a major drawback since they reflect the risk at the project level. Moreover, the classical statistical indicators provide insight into project financial risk without taking into account the influence factors or

correlations that exist between the various projects [3]. FLS have the advantage of being able to identify the financial risk for the entire portfolio of projects contained by the operational program and to contribute to the decision of management to avoid or eliminate the financial risk. The FLS input variables of this model are set according to the project particularities, on which ultimately depend the financial risk of the project, or project value (VP) and the duration of their implementation (DI). The output variable is the financial risk of the project (RF). The assessment of the financial risk of the project (as an output variable of FLS) was structured by verbal expressions (specific to fuzzy logic): high financial risk, medium financial risk and low financial risk, depending on the seriousness of the risk but also to highlight the intensity of the losses from the budget allocated to EU member states as a result of the manifestation of the financial risk event [6]. The FLS developed to identify the financial risk of the projects financed from structural funds becomes a novelty in literature but also a management tool for decision makers. With the help of FLS they can measure the financial risk for all the projects entering in the structure of the operational program. Based on these ways of measuring the financial risk, identified using FLS, corrective action can be taken for the efficient and fair presentation of structural funds for member states of the EU budget.

2 The concept of financial risk of projects financed from structural funds

The financial risk of the projects financed from structural funds is a fairly new concept both in literature and in practice. In essence the financial risk of the projects should answer the following question: "What is the financial size of the potential loss that a member state is expected to suffer, due to the implementation of projects financed from structural funds?" The project financial risk depends on a number of factors determined by the main project implementation cycle. In this category are included factors that are measured through the implementation period, resulting the physical progress of projects or the requests repayment duration. In practice, the most important factor that influences the financial risk of the project remains the physical progress [11]. The physical progress, although it seems a technical term, is most often defined as the ratio between the duration of implementation of a project under implementation cycle (D_{ic}) and the duration of implementation actually achieved (D_r) according to a relation of the form:

$$P_f = \frac{D_r}{D_{ic}} \times 100 \tag{1}$$

The achieved physical progress can record higher values than those set, for a project suitable for the implementation cycle, situation in which the financial risk of the project is small or can record lower values, where the financial risk of the project increases. The financial dimension of a project risk therefore depends on the value of the project (V_p), on the physical progress established under implementation cycle (p_{fc}) and on the actual physical progress achieved (p_r) after a relationship of the form:

$$R_f = V_p(p_{fc} - p_{fr}) \tag{2}$$

or

$$R_f = V_p \times (\frac{D_{ct}}{D_{ip}} - \frac{D_r}{D_{ip}}) \tag{3}$$

The higher is the difference between the physical progress of the project under implementation cycle (p_{fc}) and the actual physical progress achieved (P_{fr}), the higher is the risk of losing a larger

amount of the budget allocated to a project. In practice, the most common form of financial risk measurement remain in the project's value (V_p) and the deviation of the achieved implementation duration of the project D_r towards the set one based on implementation cycle (D_{ci}), after a relationship of the form:

$$P_f = V_p \times \frac{D_{ci} - D_r}{D_{ci}} \times 100 \tag{4}$$

There are projects with high financial risk, for which the difference between the two periods (according to implementation cycle and the one actually achieved) is far from average, or projects with low financial risk for which the same difference is close to average. The average difference between the two periods is determined for all projects that are part of the operational program. To appreciate the intensity of the financial risk of a project it can be used the standard deviation or variance [13], characterizing the removal from the average financial risk of a project after a relationship of the form:

$$R_f = \sqrt{\frac{1}{N}} \sum_{i=1}^{n} R_{fi} - \overline{R_f} \tag{5}$$

Depending on the value of this statistical indicator, the intensity of the financial risk is estimated. The higher is the percentage value of this indicator, the higher is the financial risk of lossing the budget allocated to projects. The indicator, thus, doesn't provide a complete picture of risk, as it is determined based on historical values recorded in previous periods of time [2]. In FLS , the financial risk of the projects will be divided into three categories (large, medium and small), considered as an output of the system. In essence, the financial risk of a project, regardless its quantification and intensity, depends on a number of factors that will transform the input for FLS, as follows [10]:

1. The value of projects, that influences the size of the financial risk: The rule is simple: The higher is the project value and the deviation from the implementation period, the higher is the financial risk

2. The project implementation duration, which is determined according to the intensity of risk: The implementation period is known in the literature as the duration necessary to implement a project, so that it can be achieved its project objectives and outcome indicators. The further the implementation period of this project removes from the one set according to the implementation period, the higher is the risk that a part of the budget will be lost.The project value and its implementation duration, as influencing factors underlying the financial risk will be considered for FLS as input variables.

3 The development of the FLS to identify financial risk of a project

Each fuzzy logic system supposes four distinct phases [14] as follows: setting the input variables and their associated fuzzy sets, the fuzzy rule base identification, the establishment of fuzzy inference operators and defuzzification. In the first stage of development of the fuzzy logic system were established the input variables mentioned above, namely: the project value (PV) and the duration of implementation (DI), while the output variable (result) is the financial risk of the project. The input variables are structured according to the size of their impact on financial risk. The project value (VP) is divided into three categories, namely: high value

projects (HVP), the average project value $VP_{\overline{m}}$, and the small projects value VP_m. At the same time, the duration of project implementation is structured according to the exceeding of the project implementation duration to that established in the implementation cycle as follows: high exceeding (between 180 days and 360 days) (DIM), average exceeding (between 90 days and 180 days) $DI_{\overline{m}}$ and low exceeding DI_m under 90 days. The financial risk of the project, as the output variable of the system, is determined by the fuzzy base rules, the fuzzy inference operators as well as by the expert assessments, that are divided into three categories namely: high financial risk (RFM), average financial risk $RF_{\overline{m}}$ and low financial risk RF_m. In order to completely define the fuzzy set, for the input variables were established the following membership functions [3]:

1. The trapezoidal membership function, for the value of projects (VP) defined according to the value types of projects

2. The triangular membership function, for the exceeding duration of the project implementation (DI) which was also structured in: high, medium and low exceeding;

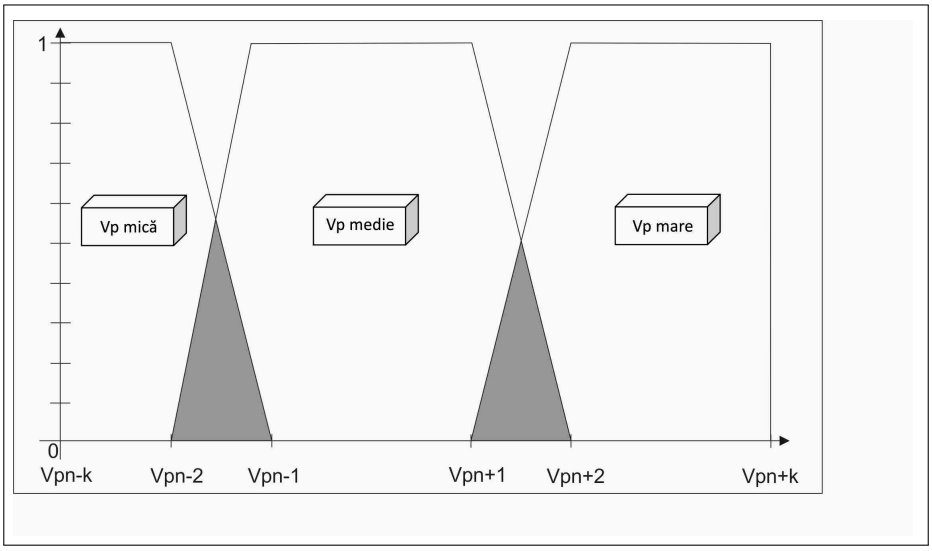


Figure 1: The trapezoidal membership function for the input variables

The fuzzy sets allow the partial membership of elements, the fuzzy membership degree being able to take any value from 0 (not belonging) to 1 (full membership). For example, the trapezoidal membership's function for $VP \in [VP_{n-2}, VP_{n+2}]$, and for project with medium values will be expressed as:

$$\mu_{VPM}(VP) = \begin{cases} 0 & \text{if } VP_{n-2} < VP \text{ and } VP > VP_{n+2} \\ \frac{VP - VP_{n-2}}{VP_{n-1} - VP_{n-2}} & \text{if } VP_{n-2} \leq VP \leq VP_{n-1} \\ 1 & \text{if } VP_{n-1} \leq VP \leq VP_{n+1} \\ \frac{VP_{n+2} - VP}{VP_{n+2} - VP_{n+1}} & \text{if } VP_{n+2} \leq VP \leq VP_{n+1} \end{cases}$$

Similarly, the trapezoidal membership function for the interval $VP \in [VP_{n-k}, VP_{n-1}]$, which corresponds to the values of small projects, which will be expressed as follows:

$$\mu_{VPm}(VP) = \begin{cases} 0 & \text{if } VP < VP_{n-k} \text{ and } VP > VP_{n-1} \\ \frac{VP_{n-1}-VP}{VP_{n-1}-VP_{n-2}} & \text{if } VP_{n-2} \leq VP \leq VP_{n-1} \\ 1 & \text{if } VP_{n-k} \leq VP \leq VP_{n-2} \end{cases}$$

The triangular membership function for the input variable, the exceeding of the duration of implementation (DI) of the project is represented in figure no.2. The triangular membership function can be expressed, for example, for the low exceeding of the duration of implementation by the relationship, ($DI \in [DI_{n-k}, DI_{n-1}]$):

$$\mu_{DIm}(DI) = \begin{cases} 0 & \text{if } DI_{n-k} < DI \text{ and } DI > DI_{n-1} \\ \frac{DI-DI_{n-k}}{DI_{n-2}-DI_{n-k}} & \text{if } DI_{n-k} \leq DI \leq DI_{n-2} \\ \frac{DI_{n-1}-DI}{DI_{n-1}-DI_{n-2}} & \text{if } DI_{n-2} \leq DI \leq DI_{n-1} \end{cases}$$

Similarly, the triangular membership function for the input variable, the duration of the implementation of the project on the interval $[DI_{n-2}, DI_{n+2}]$], can be expressed as:

$$\mu_{DI\overline{m}}(DI) = \begin{cases} 0 & \text{if } DI_{n-2} < DI \text{ and } DI > DI_{n+2} \\ \frac{DI-DI_{n-2}}{\dot{D}I_{n-2}-DI_{n-2}} & \text{if } DI_{n-2} \leq DI \leq \dot{D}I_{n-2} \\ \frac{DI_{n+2}-DI}{DI_{n+2}-\dot{D}I_{n-2}} & \text{if } \dot{D}I_{n-2} \leq DI \leq DI_{n+2} \end{cases}$$

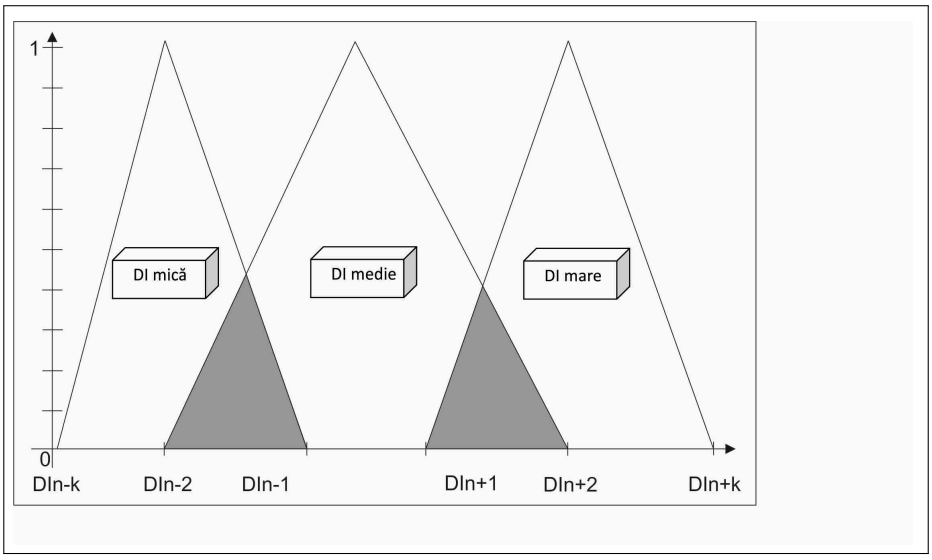


Figure 2: The triangular membership function for input variable of the system - The duration of implementation (DI)

For the FLS were identified the input variables of the system and their membership functions according to fuzzy rules, to which a fuzzy set is completely determined by the set of ordered pairs [7, 8]:

$$A = \{(x, \mu_A(x))/x \in X\} \tag{6}$$

In the second stage of the FLS are set the fuzzy rules base. The specific of these rules is that based on two conditions, namely "if" and ‘"then", are established with the help of experts the

influence factors of the financial risk. The number of fuzzy rules base will be equal to $3^2 = 9$ and the financial risk will be divided into three risk classes. The fuzzy rules base for financial risk related to projects financed from structural funds will have the following form:

Rule 1: If the project value is large (VPM) and the exceeding of the duration of the implementation is high (DIM), then the project financial risk is high (RFM);

Rule 2: If the project value is average $VP_{\overline{m}}$ and the exceeding the implementation is high (DIM), the financial risk is high (RFM);

Rule 3: If the project value is small $VP_{\overline{m}}$ and the exceeding of the duration of the implementation is high (DIM), then the financial risk is medium $RF_{\overline{m}}$;

Rule 4: If the project value is large (VPM) and the exceeding of the duration of the implementation is average $DI_{\overline{m}}$, then the financial risk is high (RFM);

Rule 5: If the project value is average $VP_{\overline{m}}$ and the exceeding of the duration of the implementation is average $DI_{\overline{m}}$, then the financial risk is medium $RF_{\overline{m}}$;

Rule 6: If the project value is small VP_m and the exceeding of the duration of the implementation is average $DI_{\overline{m}}$, then the financial risk is small RF_m;

Rule 7: If the project value is large (VPM) and the exceeding of the duration of the implementation is small DI_m, then the financial risk is medium $RF_{\overline{m}}$;

Rule 8: If the project value is average $VP_{\overline{m}}$ and the exceeding of the duration of the implementation is small DI_m, then the financial risk is medium $RF_{\overline{m}}$;

Rule 9: If the project value is small VP_m and and the exceeding of the duration of the implementation is small DI_m, then the financial risk of the project is small RF_m;

The fuzzy rules base for FLS that targets the projects financial risk, aim to capture the best way in which financial risk occurs when there is a change in the value of projects and concomitant a change in terms of exceeding the project implementation duration. Depending on the intensity of these changes, the risk of project budget loss may be, as mentioned above: large, medium or small. On the third stage were applied the fuzzy inference operators on the rules basis generated in the second stage [4]. As shown, the fuzzy rules base is connected by "AND" which means that the operator inference for the rules base is minimum. For each of the previous defined fuzzy rules, is established the degree of membership of the output variable (RF). Therefore will result:

For rule 1: $\mu_{RFM,1} = min[\mu_{VPM}(VP), \mu_{DIM}(DI)]$
For rule 2: $\mu_{RFM,2} = min[\mu_{VP\overline{m}}(VP), \mu_{DIM}(DI)]$
For rule 3: $\mu_{RFm,3} = min[\mu_{VPm}(VP), \mu_{DIM}(DI)]$
For rule 4: $\mu_{RFM,4} = min[\mu_{VPM}(VP), \mu_{DI\overline{m}}(DI)]$
For rule 5: $\mu_{RF\overline{m},5} = min[\mu_{VP\overline{m}}(VP), \mu_{DI\overline{m}}(DI)]$
For rule 6: $\mu_{RFm,6} = min[\mu_{VPm}(VP), \mu_{DI\overline{m}}(DI)]$
For rule 7: $\mu_{RF\overline{m},7} = min[\mu_{VPM}(VP), \mu_{DIM}(DI)]$
For rule 8: $\mu_{RF\overline{m},8} = min[\mu_{VP\overline{m}}(VP), \mu_{DIm}(DI)]$
For rule 9: $\mu_{RFm,9} = min[\mu_{VPm}(VP), \mu_{DIm}(DI)]$

From the rules analysis is shown that the affiliation of the system financial risk in a fuzzy set can be from one or more fuzzy rules which are likely to result in different degrees of belonging to the same fuzzy set. But it takes a single degree of belonging and in order to establish it, is applied a fuzzy controller MAX corresponding to the reunion of the fuzzy sets. Under these conditions will result [9]:

$\mu_{RFM} = max[\mu_{RFM,1}, \mu_{RFM,2}, \mu_{RFM,4}]$
$\mu_{RF\overline{m}} = max[\mu_{RF\overline{m},5}, \mu_{RF\overline{m},7}, \mu_{RF\overline{m},8}]$
$\mu_{RFm} = max[\mu_{RFm,3}, \mu_{RFm,6}, \mu_{RFm,9}]$

In this stage are obtained the solutions of the fuzzy rules, without a certain amount of input variables in the system (VP_i, DI_i) to be determined the intensity of the financial risk by applying all rules of financial risk in the fuzzy base. It is therefore necessary to identify these solutions at

the last stage of FLS, namely defuzzification. In the last stage of defuzzification is extracted a deterministic scalar value, from the fuzzy information which is associated to the output variable, the essence of which is to provide more explicit the best value of the output variable [1]. Each of the result in the third stage of the FLS will be used to determine the surface area (S_i) bounded by the parallel to the horizontal axis, taken through the point that determines the size of the output variable, the horizontal axis (O_x) and the graphic of the function associated to output variables. For a given value of the project (VP_i) and some exceeding of the project implementation duration (DP_i) would result that from the original surface (S_i) only a certain percentage (p) is the result that will be taken into account to determine the final amount of financial risk [5]. The conversion of the fuzzy result in a real number value is done by determining the center of gravity of the surface obtained by aggregating the proportion (p) of the initial areas for each graphical input variables as follows:

$$Z = \frac{\sum_{i=1}^{9} \mu_{VP,DI}(VP_i, DP_i) \times VP_i/DP_i}{\sum_{i=1}^{9} \mu_{VP,DI}(VP_i, DI_i)} \tag{7}$$

The result (S) of equation 7 undergo a conversion from area into a real value through the centroid method, which consists in determining the numerical value (z) through which the perpendicular traced to the horizontal axis divides the S area into two equal parts by a relationship of the form:

$$S = \bigcup_{i=1}^{9} \%pS_i \tag{8}$$

The obtained numerical Z value represents the size of the financial risk of a project that has a certain value and a certain level recorded for the (exceeding) implementation duration of the project. The higher this value is, the higher is the probability of losing a part of the project budget.

4 The FLS simulation for the financial risk of projects financed from structural funds

The developed fuzzy logic system was simulated using MATLAB programming language, taking into account the following assumptions, namely:

1. The input variable - the value of projects was divided into three classes: small projects (between 0 and 350 million UM), average project (between 250 million and 750 million U.M.) and large projects (between 650 million UM and 1,000 mil);

2. The input variable - the exceeding of the project implementation duration was also divided into three classes, namely: low exceeding (between 0 and 90 days), average exceeding (between 60 to 180 days) and high exceeding (between 150 days and 270 days).

In the first FLS stage, were established the input variables and their membership functions. Thus, for the input variable, the project value (PV), was stated the following fuzzy set (trapezoidal membership function) as depicted in Figure 3.

For the input variable, the exceeding of the project implementation duration (DI), resulted the following fuzzy sets (using the trapezoidal membership function) as depicted in Figure 7.

For the output variable, the financial risk of the projects, were established three risk classes, using the triangular membership function as follows: low financial risk for values between 0 and

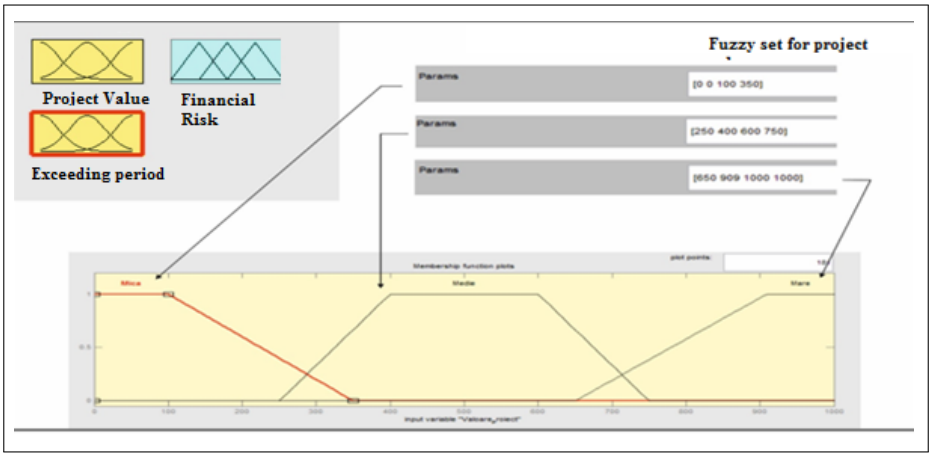


Figure 3: The fuzzy set for the input variable - The project value

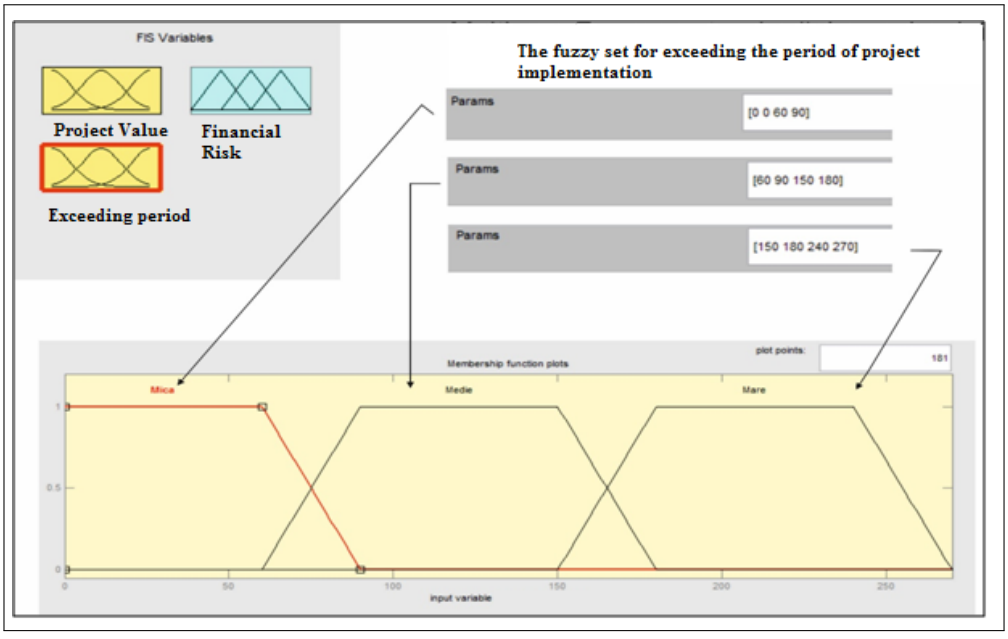


Figure 4: The fuzzy set for the input variable - The implementation duration

3, average financial risk for values between 2 and 6 and greater financial risk for values between 5 and 10. The resulted fuzzy set for the output variable is represented in Figure 5.

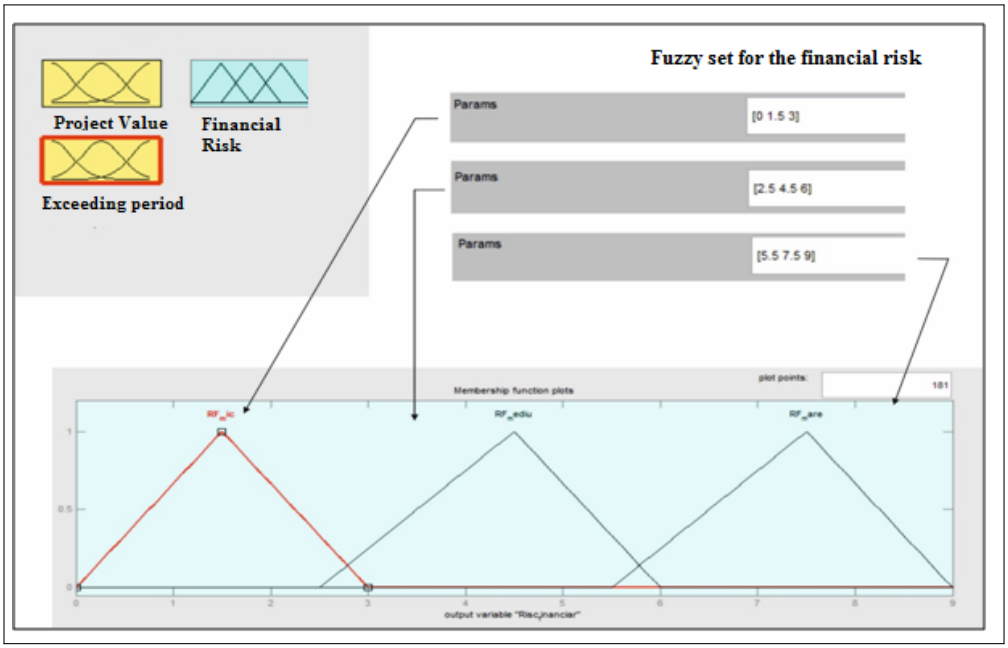


Figure 5: The fuzzy set for the output variable - The financial risk of the projects

After continuing the simulation of FLS, for the financial risk of the projects, were established 9 fuzzy rules base according to the developed and were introduced in the program shown in Figure 6.

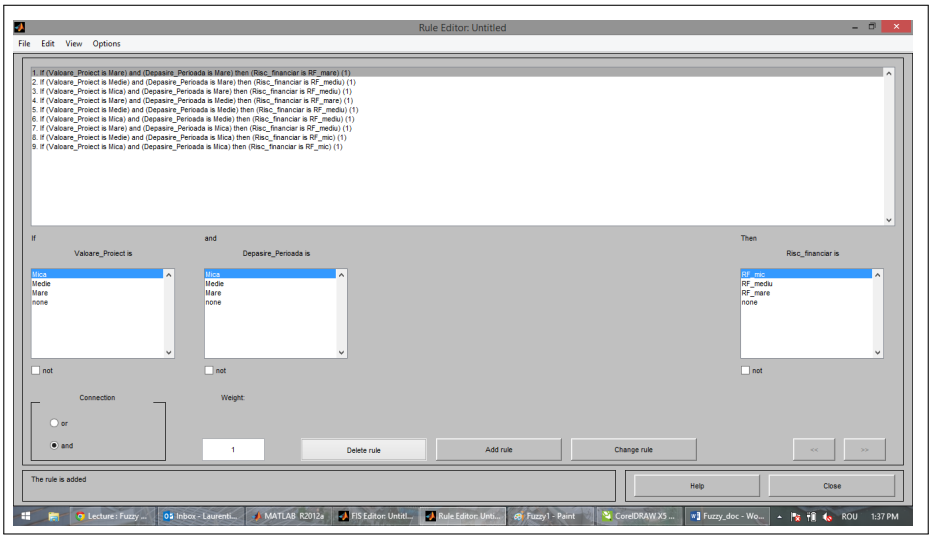


Figure 6: The fuzzy rules base for the financial risk in Matlab

Subsequently were applied the inference operators on fuzzy rules base and the results were presented in Figure 7.

The obtained results are for a project with a value of 500 million U.M., 135-days exceeding duration of the project implementation and a financial risk value of 4.33. The FLS simulation for the financial risk was further carried out for different values of input variables in order

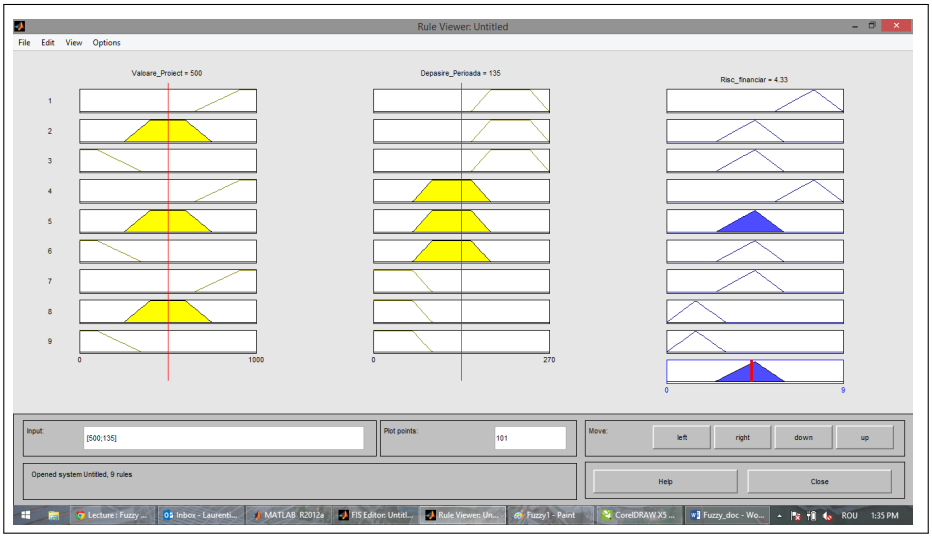


Figure 7: Results obtained by applying the inference operators on fuzzy rules base

to identify the developments that the financial risk has when changes occur in values and in exceeding duration of the project implementation. The results were obtained in Figure 8.

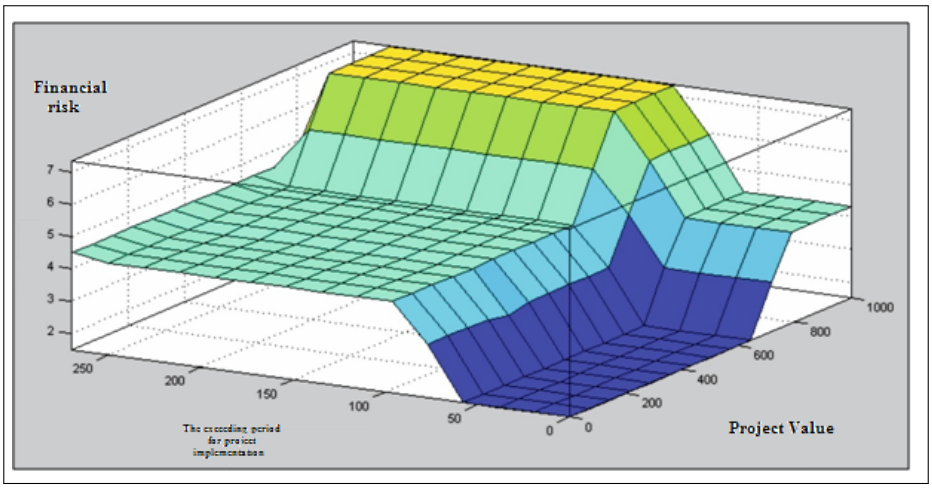


Figure 8: The evolution of financial risk for projects based on input variables

The simulation results are:

1. For project values between 0 and 600 million and an exceeding duration of the project implementation of 55 days, the financial risk tends to zero;

2. For project values between 600 million and 800 million and an exceeding duration of the project implementation between 55 and 100 days, the financial risk is increased from 0 to about 4.5 (and is considered a medium risk);

3. For project values between 800 million and 1,000 million and an exceeding duration of the project implementation between 100 and 270 days, the financial risk is increased from 4.5 to the maximum 9 (being considered a high risk);

5 Conclusion

The computational intelligence is an area of great interest for both specialists in computer science and finance. This is because often are abandoned the statistical concepts and methods that didn't characterize the phenomena and economic processes and are used specific computational intelligence methods that adapt quite well the dynamics of the phenomena studied. The fuzzy logic system developed in this paper, as part of computational intelligence, studies the behavior of financial risk of the projects financed from structural funds when there is a change in the value of projects or in the durations of implementation. This financial risk was defined as the risk of losing a part of the project budget, when there are exceedings in the duration of implementation, based on their implementation cycles. The rules of the fuzzy base were defined according to the impact that the system input variables have on financial risk. The fuzzy inference operators were applied on the basis rules to determine the membership in fuzzy output variable. With the help of defuzzification was ensured the convertion of the fuzzy values in numeric values to determine the size of financial risk for different values of the input variables. Following the simulation for the output variable (the financial risk of projects), were reached the following conclusions:

1. There are situations where the financial risk is zero, or almost zero, for different values of input variables in the system;

2. The financial risk of projects increases as changes occur in the value of projects and in the duration of implementation. This increase in financial risk value becomes in time proportional to changes in the input variables in the system;

3. The financial risk is maximum, when the input variables in the system approach the maximum values that they may register;

The developed fuzzy logic system is a management tool for decision making in the structural funds, under which can be taken measures to avoid or reduce the financial risk, especially for the early identification of the emerging financial risk in the portfolio of projects falling within the structure of the operational program. This early identification of financial risk can be very useful in structural funds management, for taking the necessary actions to avoid this category of risk [12].

The financial dimension of risk can be determined by calculating the financial risk using the formula 4 for each project, where FLS indicates values above the minimum for the financial risk. The value obtained is the value of projects that may be at risk of losing resources that are allocated from the Structural Funds through operational programs. The fuzzy logic can be also developed on classes of projects, through clustering technique to identify financial risk for each class of projects which will be useful in the future because each operational program includes an amount of projects.

Bibliography

[1] A. Altrock, Fuzzy logic and Neuro Fuzzy Logic Applications Explained (1995); *Prentice Hall*, Englewood Cliffsm.

[2] I.A. Bradea, C. Delcea, R.M. Paun (2014); Managing and Controlling the KRIs in Hospitals *Proceedings of 24rd IBIMA Conference: Crafting Global Competitive Economies: 2020 Vision Strategic Planning & Smart Implementation*, Italy, ISBN: 978-0-9860419-3-8, 1824-1830.

[3] I.A. Bradea (2014); Risks in Hospitals. Assessment and Management *The Romanian Economic Journal*, ISSN: 2286-2056, 54(XVII): 25-37.

[4] R. Fuller (2000); Introduction to Neurofuzzy Systems, *Advances an Inteligent and Soft Computing*, vol 2, ISBN 978-3-7908-1256-5.

[5] C. Kahraman, I. Kaya (2010); A Fuzzy Multicriteria Methodology for selection Among Energy Alternatives, *IExpert Systems with Applications*, 37(9): 6270-6281.

[6] S.M. Mousave, F. Joloi, R. Tavakkoli-Moghaddam (2013); A Fuzzy Stochastic Multi-Attribute Group Decision-Making Approach for Selection Problems, *Group Decision and Negotiation*, 22(2): 207-233.

[7] S. Nadaban (2015); Fuzzy Euclidean Normed Spaces for Data Mining Applications, *International Journal of Computers Communications & Control*, 10(1): 70-77.

[8] S. Nadaban, I. Dzitac (2014); Atomic Decompositions of Fuzzy Normed Linear Spaces for Wavelet Applications, *Informatica*, http://dx.doi.org/10.15388/Informatica.2014.33, 25(4): 643-662.

[9] A. Nieto-Morote, F. Ruz-Vila (2011); A Fuzzy Approach to Construction project Risk Assessement, *International Journal of Project Management*, 29(2): 220-231.

[10] E. Scarlat, N. Chiriţă, I.A. Bradea (2012); Indicators and Metrics Used in the Enterprise Risk Management (ERM), *Economic Computation and Economic Cybernetics Studies and Research Journal*, 4(46):5-18.

[11] M.L. Tseng (2010); Implementation and Performance Evaluation Using the Fuzzy network Balanced Scorecard, *Computers & Education*, 55(1): 188-201.

[12] Y.L. Xu, J.F.Y. Yeung, A.P.C. Chan, D.W.N. Chan, S.Q. Wang, Y.L. Ke (2010); Developing a Risk Assessement Model for PPP project in China - A Fuzzy Synthetic Evaluation, *Automation in Construction*, 19(7): 9293-943.

[13] L.A. Zadeh (1992); Fuzzy logic and the calculus of fuzzy if-then rules, *Proceedings of the 22nd Intl. Symp. on Multiple-Valued Logic*, Los Alamitos, CA: IEEE Computer Society Press, 530-561.

[14] L.A. Zadeh (1996); Fuzzy logic - computing with words, *IEEE Trans. Fuzzy Systems*, 4(2):103-111.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):492-499, August, 2015.

Restructured Ant Colony Optimization Routing Protocol for Next Generation Network

B. Chandramohan

Dr. B. Chandramohan
Kanchi Pallavan Engg College,
Anna University, Chennai, India
abc.phd@hotmail.com

Abstract: Wireless network is a major research domain in the past few decades. Wireless network evolves in many forms like cellular communication, ad hoc network, vehicular network, mesh network and sensor network. Next generation network is a recent cellular communication which provides heterogeneous connectivity on cellular communication. The routing in next generation wireless networks is an important research issue which requires many constraints than wired networks. Hence, Ant Colony Optimization (ACO) is applied in this paper for routing in heterogeneous next generation wireless network. The ACO is a swarm intelligence technique which applied for many engineering applications. ACO is an optimal technique for routing and travelling salesman problem. This paper proposed Restructured ACO which contains additional data structures for reducing packet loss and latency. Therefore, the proposed RACO provides higher throughput.
Keywords: Wireless Network, Next Generation Network, Routing, Swarm Intelligence, Ant Colony Optimization.

1 Introduction

Swarm intelligence is a new discipline of study that contains a relatively optimal approach for problem solving, which is the imitation inspired from the social behaviours of insects and of other animals. ACO, artificial bee colony algorithms and fire fly algorithm are few techniques in swarm intelligence. The ACO is an optimization technique which is widely applied for a variety of optimization problems in various engineering field of studies.

The few application of ACO in the recent year are, Job Scheduling (Li-Ning et al 2010), Project Scheduling (Wang Chen et al 2010), Production management and maintenance scheduling (Osama et al 2005), Cash Flow Management (Wei-Neng et al 2010), Manpower Scheduling and management (Hsin-Yun et al 2010), Travelling Salesman Problem (Xiao-ming et al 2010), Clustering and set partitioning (Ali and Babak 2010), Pattern Recognition (Zhiding et al 2010). Dorigo et al (1996, 1997, 2004) proposed Ant System, Ant Colony System and Ant Net, which are the significant implementation of ACO. Dorigo et al (1996) applied the simple probability rule and Dorigo and Luca (1997) applied the state transition rule for the decision model. Dorigo and Stutzle (2004) redefined the pheromone update policy of ACO, and the term argentine ant is replaced with forward ant.

In the network routing, Ant-Net Routing using ACO provides a optimality due to its real time computation and less control overhead. Kwang and Weng (2003) comparing all routing algorithms with ACO, concludes that ants are relatively small, so it can be piggybacked in data packets and more frequent transmission of ants may be possible in order to provide updates of routing information for solving link failures. Hence, using ACO for routing in dynamic network seems to be appropriate. Routing in ACO is achieved by transmitting ants rather than routing tables or by flooding LSPs. Even though it is noted that the size of an ant may vary in different systems or implementations, depending on their functions and applications, in general, the size of ants is relatively small, in the order of 6 bytes.

Comparison of network performance on OSPF and ACO is available in Chandramohan et al (2007). The ACO Based Redundant Link Avoidance Algorithm is proposed in our earlier paper (Chandramohan et al, 2010). Detailed study on ACO in various engineering domain is available in Chandramohan et al (2011, 2012). This paper further explains the data structure of proposed Restructured ACO (RACO) in section 2, the route discovery and route maintenance of proposed RACO in section 3, and routing decision model of RACO in section 4. and the results and performance analysis in section 5.

2 Data Structure of Proposed RACO

The existing ACO has two data structures, namely, pheromone matrix and traffic model. The proposed RACO restructures the ACO by introducing two more data structures called traffic model II and availability model. These two data structures are used for identifying congestion free and stable routes. In addition to the extended data structures, the routing decision of proposed RACO is based on compound probability rule. The compound probability rule is the mean value of the probability of existing random probability rule and the two additional probability rules proposed in this work. The proposed two additional probability rules are namely probability based on response complexity and probability based on packet loss complexity. In addition to this compound probability rule, the priority model based on availability model is defined for single path routing in the wireless routing.

Hence, there are four data structures used in the proposed RACO, namely 1) Pheromone Matrix Model, 2) Traffic Model I, which is used as Round Trip Time (RTT) traffic model, 3) Traffic Model II, which is used as Packet Loss traffic model and 4) Availability Model.

3 Route Discovery and Route Maintenance in RACO

Every source node in the network, in a regular interval (t), generates forward ants (FA) and the FA is circulated in the network for searching the destination. Destinations are locally selected according to the data traffic patterns generated by the local workload. Both intermediate and source nodes forward the FA in the same way. The FA carries the path source address, the destination address, the intermediate node Identification, and the path information. The FA generation rate can be a function of network dynamics, data rate, time, etc. The FA moves in the network searching for the destination using the probability routing table of intermediate nodes.

While moving, the FA collects the information about the time length or trip time, the congestion status, and the node identifier of the intermediate nodes. A sufficient number of the ants visit the neighbour corresponding to the highest probability in the routing table. However, a number of the FA have a probability to visit other nodes and other paths still have a probability to be visited. This will increase the number of the FA visiting nodes in the region around the best path. In addition, it allows a fair number of FA to visit other regions in the network. Unlike flooding, a FA will be forwarded to only one neighbour.

When a FA reaches its destination, the information carried by this FA path will be graded. Then, the FA will be killed and a Backward Ant (BA) will be generated in the destination. The BA carries its corresponding FAs path grade and paths intermediate nodes ID ant it will be send back to the source node by following the reverse path of its corresponding FA. As the BA moves in the reverse path, the intermediate nodes modify their four data structures based on the path grade carried by the BA and accordingly update their probability routing tables.

Finally, the source node receives the BA, updates its tables, and kills the BA. FA shares the

same queues as data packets, so that they experience the same traffic load. The BA takes the same path as the concern FA travelled, but in opposite direction. BA do not share the same link queues as data packets (like FA), they use the higher-priority queues reserved for routing packets, since the only task of BA is to quickly propagate to the pheromone matrices (the information accumulated by the FA).

Ants adapt their exploration activity in this way to the varying data traffic distribution. While travelling towards their destination nodes, the FA keep memory of their paths and of the traffic conditions found. The identifier of every visited node and the time elapsed since the launching time to arrive at this node is stored in a memory stack S(i), where, i - is the node identifier. Further the ant builds a path by performing the following steps:

Step 1 (Distribution of Forward Ant): At each source node(s), each FA headed toward the destination node(d) by selecting the intermediate node(i), choosing among its neighbour(j) which is not visited already. The neighbours node(j) is selected with a probability P(i,j,d) computed as the normalised sum of the pheromone T(i,j,d) with a heuristic value N(i,j) taking into account the state (the length) of the jth link queue of the current node I is

$$P(i,j,d) = \frac{T(i,j,d) + \alpha * N(i,j)}{1 + \alpha(|N(i)| - 1)} \tag{1}$$

The heuristic value N(i,j) is a normalised value function (0, 1) of the length of the queue q(i,j) between the node i and its neighbour j is

$$N(i,j) = 1 - \frac{q(i,j)}{\sum_{i=1}^{N} q(i,j)} \tag{2}$$

The value of α determines the importance of the heuristic value with respect to the pheromone values stored in the pheromone matrix (T). The value N(i,j) reflects the instantaneous state of the nodes queues and assuming that the queues consuming process is almost stationary or slowly varying, N(i,j) gives a quantitative measure associated with the queue waiting time. The pheromone values, on the other hand, are the outcome of a continual learning process. This learning process will register both the current and the past status of the whole network. The two components of the ant decision system, T and N, used for the decision system will act both the combination of long-term learning process and an instantaneous heuristic prediction.

Step 2 (Avoid Cyclic Routing): If an ant returns to an already visited node, the cyclic nodes are removed and all the memory about them are deleted. This process helps the system to avoid the count-to-infinity problems.

Step 3 (Generate Backward Ant):
* Condition1 [Generate BA and delete FA] - When the destination node(d) is reached, the FA generates another ant called the backward ant and transferred the information stored in the memory of FA to BA. Then the FA is to be deleted.
* Condition2 [Delete FA on lifetime Basis] - The FA is deleted if its life time becomes greater than a value of MaxLife before it reaches the destination node, which is similar to TTL (Time to Live) in the TCP.

Step 4 (Flow of BA): The flow of FA and BA have some distinctions. The BA takes the same path as the concern FA travelled, but in the opposite direction. The FA share the same link queues as data packets, but the BA do not share the same link queues as data packets, instead BA uses the higher-priority queues reserved for routing packets. Because the task of

BA is to quickly propagate to the pheromone matrices. The pheromone matrix is updated by incrementing the pheromone T(i,f,d) and decrementing the other pheromones T(i,j,d), where, i-represents the node ID, d-is the destination node, f-is the current neighbour and j-is the existing neighbour, where j is not equal to f. The trip time is a good decision parameter because it indicates the appropriateness of the physical and logical condition of the path, which includes the number of hops, transmission capacity of the concerned link, processing time and queuing delay. A path with a high trip time also considered as a good path, if its trip time is significantly lower than the other trip time.

The traffic model is updated with the values stored in the BAs memory. The time elapsed to arrive to the destination by the FA starting from the current node is used to update, according to the following equations:

$$\mu_i = \mu_i + \delta(RTT(i,d) - \mu_i) \tag{3}$$

$$\beta_i^2 = \beta_i^2 + \delta(RTT(i,d) - \mu_i)^2 \tag{4}$$

where RTT(i,d) is the new observed trip time from node i to the destination d. μ is the sample mean of the traffic model and β^2 is the variance of the traffic model. δ is the representation of the weighs of the number of most recent samples that will really affect the average.

The value of δ is calculated as $\delta = 0.1$, assuming that the latest fifty observations really influence the estimate (approximately), i.e. (5/50). Suppose 100 latest observations really influence the estimate, then $\delta = 5 \ / \ 100 = 0.05$. Node i, waits for a specific predefined time unit (maximum allowable trip time) for the BA of each sending FA. If the BA is not reached within the time period, then the node i considered the ant is destroyed in the intermediate node due to heavy congestion or by wrong propagation in the path.

For every interval (t), the number of packet loss is calculated and this information is stored in the traffic model II.

$$PL_{new} = PL_{old} + PL_{observed} - \mu * PL_{old} \tag{5}$$

where PL_{new} is the new value for packet loss, PL_{old} is the existing value of packet loss, $PL_{observed}$ is the currently calculated value of packet loss, and μ is the standard deviation of existing values. This is an optional model that used only for wireless environment. This availability model also may be opted for wired environment in order to test the frequent link failures. This model stores the information of the availability of the next hop based on its battery power and / or the mobility. When a BA reaches node i, then the information stored regarding the mobility or link failure in the memory is collected and this model is updated. The value of the availability model is between (0,1). The value 0 represents the next hop is unavailable and the value 1- represents the next hop is available. If the existing value is 1, the new value is 0, and then the new value is updated. If the existing value is 0 (or 0 to 1), the new value is 1, and then the current value is defined as, current value = observed value $-(\frac{1}{1+existingvalue})$.

4 Routing Decision Model

The routing decision model computes and identifies the optimal path from the information stored in the data structures. In the RACO, the routing model applies the compound and priority based routing model for routing decision.
Response Complexity (C_r) of particular path is calculated using

$$C_r = \frac{rtt_s - rtt_a}{rtt_s} \tag{6}$$

where, rtt_s is standard RTT (Round Trip Time), and rtt_a is actual RTT. The standard RTT is defined based on the applications, for data communication the standard RTT is defined as 150ms and for the multimedia communication the same is defined as 100ms in order to avoid problems like jitter.
Probability based on Response Complexity (P_c) is

$$P_c = \frac{C_r}{\sum C_r} \tag{7}$$

Similarly, the Packet Loss Complexity (C_{pl}) of particular path is computed using

$$C_{pl} = \frac{C_{pl}}{\sum C_{pl}} \tag{8}$$

and the Probability based on Packet Loss Complexity (P_{pl}) is

$$P_{pl} = \frac{P_{pl}}{\sum P_{pl}} \tag{9}$$

The Random proportional rule is applied for Pheromone density ($P_{T,j}$) is

$$P_{(T,j)} = \frac{(T_j + k)^h}{\sum_{i=1}^{n} (T_i + k)^h} \tag{10}$$

where k and h are constant, k = 0,1 and h = 0,1. From the above, the compound probability is obtained as

$$P = \frac{P_c + P_{pl} + P_{(T,j)}}{3} \tag{11}$$

For Multi Path Routing, data is propagated via all available paths based on compound probability. For Single Path Routing, the first priority path is selected and used for communication.

4.1 Pseudo Code for Route Discovery phase of Proposed RACO

A. Procedure RACO (t, t_{end}, t1)
Inputs:
t – current time
t_{end} – Time length
t1 – Time interval between ants generation
//Initialization Phase
for-each i in C do
PM – initialize pheromone matrix
TM1 – initialize traffic model I
TM2 – initialize traffic model II
AM – initialize Availability model
R – Initialize Routing Table
//Generate and Propagate Forward Ant
while $t \leq t_{end}$ do
in-parallel
if (t mod t1) = 0 then
$destination = select(distribution_equation)$
LaunchForwardAnt(Source, Destination)

end-if
for-each (ActiveForwardAnt (Source, Current, Destination)) do
$while(current \neq destination)$ do
$next_{hop} = SelectLink(current, destination, link_{queues}, T)$
$PutAntOnLinkQueue(current, next_{hop})$
$WaitOnDataLinkQueue(current, next_{hop})$
$CrossLink(current, next_{hop})$
$Memorize(current, next_{hop})$
$current = next_{hop}$
end-while
//Generate and Propagate Backward Ant
LaunchBackwardAnt(Source, Destination)
end-for each
for-each (ActiveBackwardAnt(Source, Current, Destination)) do
$while(current \neq destination)$ do
$next_{hop} = PopMemory$
$WaitOnHighPriorityLinkQueue(current, next_{hop})$
$CrossLink(current, next_{hop})$
from=current
$current = next_hop$
UpdateLocalTrafficModel(PM,TM,AM,current,from,source,memory)
GetNewPheromone (PM,current,from,source,memory)
UpdateLocalRoutingTable (PM,TM,AM,R,source)
end-while
end-for each
$end - in_{parallel}$
end-while
end-for each
end-procedure

5 Conclusion

The proposed RACO is implemented in Network Simulator 2 (NS2). The simulation performed for 10 seconds and the performances are recorded. The test are carried out in a variety of design and topology of network which include wireless network, and HNGN. Also the tests are carried out in various number of nodes; and on various load condition. The Routing Packet Size, Route Discovery Time, Packet Delivery Ratio and Network efficiency are shown in tables 1, 2, 3 and 4, respectively.

Table 1: Routing Packet Size (in Bytes)

DSR	DSDV	AODV	ACO	RACO
60	44	44	16	32

Table 2: Route Discovery Time (in ms)

DSR	DSDV	AODV	ACO	RACO
420	4300	3200	216	182

Table 3: Packet Delivery Ratio (in percentage)

No of Nodes	DSR	DSDV	AODV	ACO	RACO
100	87	89	92	96	100
200	85	88	91	95	100
500	79	84	87	91	93
1000	79	82	87	89	92
1500	76	81	85	87	91
2000	76	81	85	86	91

Table 4: Network Efficiency (in percentage)

No of Nodes	DSR	DSDV	AODV	ACO	RACO
100	71	73	88	90	94
200	69	71	85	88	90
500	67	69	83	86	88
1000	66	68	81	83	86
1500	64	66	74	74	83
2000	62	64	74	74	79

Bibliography

[1] Ali M. and Babak (2010); A. A new clustering algorithm based on hybrid global optimization based on a dynamical systems approach algorithm, *Expert Systems with Applications*, 37: 5645-5652.

[2] Chandramohan, B., Prasanna Kumar P, Anantha Venkata Ramana, Sridharan D (2007); Real time routing protocol (Antnet) using ACO and performance comparison with OSPF, *IEEE Int. Conf. on Emerging Trends in High Performance Architecture Algorithms and Computing*, 47-53.

[3] Chandramohan, B. and Baskaran, R.(2010); Improving Network Performance using ACO Based Redundant Link Avoidance Algorithm, *International Journal of Computer Science Issues*, 7(3): 27-35.

[4] Chandramohan, B. and Baskaran, R. (2011); Survey on Recent Research and Implementation of Ant Colony Optimization in Various Engineering Applications, *International Journal in Computational Intelligent Systems*, 4(4): 566-582.

[5] Dorigo, M., Maniezzo, V. and Colorni, A. (1996); Ant System: Optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, 26(1): 29-41.

[6] Dorigo, M. and Luca, M.G.(1997); Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem, *IEEE Transactions on Evolutionary Computation*, 1(1): 53-66.

[7] Dorigo, M. and Stutzle, T. (2004); *Ant Colony Optimization*, MIT Press, Cambrige MA.

[8] Hsin-Yun, L., Hao-Hsi, T., Meng-Cong, Z. and Pei-Ying, L. (2010); Decision support for the maintenance management of green areas, *Expert Systems with Applications*, 37: 4479-4487.

[9] Kwang, M. S. and Weng, H.S. (2003); Ant Colony Optimization for Routing and Load-Balancing: Survey and New Directions, *IEEE Transactions on Systems, Man, and Cybernetics*, 33(5): 60-572.

[10] Li-Ning, X., Ying-Wu, C., Peng, W., Qing-Song, Z. and Jian, X.(2010); A Knowledge-Based Ant Colony Optimization for Flexible Job Shop Scheduling Problems, *Applied Soft Computing*, 10: 888-896.

[11] NS2, available online at: www.isi.edu/nsnam/ns/

[12] Osama, H.H., Tarek, N.S. and Myung, J.L. (2005); Probability Routing Algorithm for Mobile Ad Hoc Networks Resources Management, *IEEE Journal on Selected Areas in Communications*, 23(12): 2248-2259.

[13] Wang Chen, Yan-jun, S., Hong-fei, T., Xiao-ping, L. and Li-chen, H. (2010); An efficient hybrid algorithm for resource-constrained project scheduling, *Information Sciences*, 180: 1031-1039.

[14] Wei-Neng, C., Jun Zhang, Henry Shu-Hung, C., Rui-Zhang, H. and Ou Liu (2010); Optimizing Discounted Cash Flows in Project Scheduling An Ant Colony Optimization Approach, *IEEE Transactions on Systems, Man, and Cybernetics*, Part C: Applications and Reviews, DOI:10.1109/TSMCC.2009.2027335, 40(1): 64-77.

[15] Xiao-ming, Y., Sheng, L. and Yu-ming, W. (2010); Quantum Dynamic Mechanism-based Parallel Ant Colony Optimization Algorithm, *International Journal of Computational Intelligence Systems*, Suppl. 1, 101-113.

[16] Zhiding, Y., Oscar, C.A., Ruobing, Z., Weiyu, Y. and Jing, T. (2010); An adaptive unsupervised approach toward pixel clustering and color image segmentation, *Pattern Recognition*, 43: 1889-1906.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):500-507, August, 2015.

An Energy-Efficient and Routing Approach for Position Estimation using Kalman Filter Techniques in Mobile WSNs

Y. Donoso, G. A. Montoya, F. Solano

Yezid Donoso*, Germán A. Montoya
System and Computing Engineering Department
Universidad de Los Andes
Bogotá, Colombia
ydonoso@uniandes.edu.co, ga.montoya44@uniandes.edu.co
*Corresponding author: ydonoso@uniandes.edu.co

Fernando Solano
Warsaw University of Technology
Warsaw, Poland
fs@tele.pw.edu.pl

Abstract: Mobile Wireless Sensor Networks is being an attractive field due to its applicability to an increasingly amount of mobile scenarios such as wild monitoring, disaster prevention, object guidance and health monitoring. In addition, since the sensors have limited batteries, data routing has to be planned strategically in order to extend the battery lifetime as much as possible [1] [2].
In this paper, we assume GPS free sensor devices, where considering a predictive technique to estimate the sensor position in a circular trajectory scenario can be useful to know when the sensor will be as close as possible to a sink, and then, help us to reduce the energy consumption by the fact of transmitting data at a short distance respect to the sink. In this paper, we propose an predictive algorithm based on Kalman filter techniques to estimate the proper time at which the sensor is close as much as possible to a sink, in order to reduce the energy consumption in the sensor. Specifically, we propose the usage of two Kalman Filters. One Kalman Filter is used for estimating the Received Signal Strength Indicator (RSSI) level based on several control packets received at the sensor device. This RSSI estimation indicates the distance from the mobile sensor device to the sink at a given time. The second Kalman Filter, based on the outputs from the first Kalman Filter, estimates the angular velocity and the angle of the mobile sensor device at a given time. Once this information is processed, it is possible to estimate the mobile sensor position in a circular trajectory in order to determine how much close is the mobile sensor device respect to the sink. In addition, the communication channel noise may affect the packet content, generating non-accurate information measurements at the receptor. For this reason, our proposal is evaluated under different noise channel levels and compared against a traditional technique. Our predictive routing algorithm shows better results in terms of distance accuracy to the sink and energy consumption in noisy communication channels.
Keywords: MWSN, position estimator, Double Kalman filter.

1 Introduction

The advances of Wireless Sensor Networks (WSN) have allowed attaching the sensors to an entity such as an object, animal or human, to monitor a physical variable presented in its environment. However, the sensors are equipped with limited batteries whereby it is required to implement energy efficient routing techniques to extend the lifetime of the sensors. In addition, the sensors mobility can be disadvantageous for those ones that are distant from a sink, since

this indicates employing more power transmission to set a communication link, which represents more energy consumption [1] [2].

Given the scenario described above, a possible solution would consist to implement energy efficient routing techniques considering the sensor position to know accurately when the sensor is near to a sink. Some of these solutions propose the usage of sensors equipped with GPS devices, called GPS non-free approaches. However, these GPS non-free solutions have in most of cases drawbacks such as high implementation costs, delays for acquiring position information and non-accurate position information [3]. In addition, these types of solutions require an extra chip for the GPS [4], whereby more energy consumption is experimented.

In this sense, GPS free solutions can be a viable option to solve the problem described above through using predictive techniques to estimate when it is pertinent to send data packets a sink [3]. However, the prediction processes must take into account external factors such as the inherent noise introduced by the wireless channel in the environment. For this purpose, one of the most well-known and often-used tool is the Kalman Filter. It is basically a predictor-corrector type estimator that minimizes the estimated error covariance when some presumed conditions are accomplished. The Kalman Filter has been the subject of extensive research and application, for instance in the area of autonomous or assisted navigation [8]. Thanks to the advances of digital computing, currently the Filter is practical in many applications, and due to its simplicity and robustness it can be used in sensor devices since they have severe computation and memory constraints [2], [8].

In this paper, we propose an energy efficient routing algorithm based on Kalman Filtering techniques to predict a future state of the system in order to determine the accurate time to send data packets from a sensor to a sink in despite of the wireless channel noise. Specifically, we propose the usage of two Kalman Filters. The first Kalman Filter is employed for estimating the RSSI level based on several control packets received at the sensor device. This RSSI estimation indicates the distance from the mobile sensor device to the sink at a given time. The second Kalman Filter, based on the outputs from the first Kalman Filter, estimates the angular velocity and the angle of the mobile sensor device at a given time. Once this information is processed, it is possible to estimate the mobile sensor position in a circular trajectory in order to determine how much close is the mobile sensor device respect to the sink.

The remainder of the paper is organized as follows: Section II shows in detail the problem description as well as the theoretical basis and the algorithm of our solution is shown in section III. In Section IV, we present results for the predictive approach and the classical solution showing the advantages and disadvantages of them. Finally, in Section V, we show conclusions and present future directions of our work.

2 Problem Description

The scenario to be analysed is shown in the figure 1, which it is compound of at least two sinks, s_1 and s_2, and one sensor (target) t attached to an entity (human, animal or object) describing a circular trajectory. For illustrative reasons, our application case will be an athlete monitoring scenario, in which it is assumed a sensor will be attached to an athlete for sending and collection of physiological data. The scenario characteristics is described in detail below.

In the circular trajectory, the target will have an angular velocity w_k which it is assumed to be constant. However, the angular velocity and the circular trajectory are assumed no ideal such as in real life application scenarios, whereby we assume these variations follow a Gaussian distribution with mean $\mu = 0$ and with a standard deviation σ defined in detail later. This means the target angular velocity will increase or decrease according to a Gaussian distribution along the circular trajectory, and likewise, the radius r of the circular trajectory will have a little

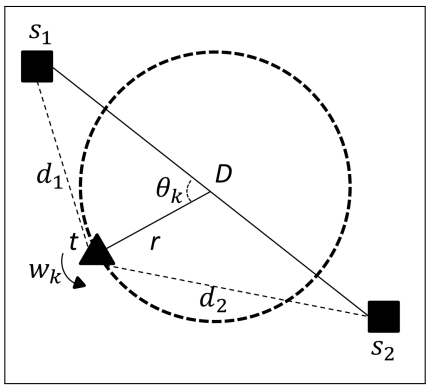


Figure 1: Application scenario.

variation in time following also a Gaussian distribution. In addition, it is assumed there are two sinks located close to the road, (s_1 and s_2) with a fixed D distance between them. Further, we assume there is always a noisy wireless link communication between sinks and the target t.

Due to the target is compound of a wireless sensor with a limited battery, it must be used efficiently in order to extend its lifetime for future usage. The target has to collect relevant information which must be sent to any sink each certain time in order to use efficiently the sensor buffer, which also is limited. For our case, we assume the best time to send data is when the target is as close as possible to a sink. Due to, for our case, it is not possible to know at any time the accurate position of the target, knowing the distance between the target and a sink could be a encouraging start to determine the target position. The distance can be inferred by extracting the RSSI value of a packet, however, the noise can disturb our measurements, and then, affecting our position estimation. As a result, our problem consists how to send the data collected by the target in an efficient energy way without knowledge of the target position in the circular trajectory, taking into account the negative noise effects caused by the wireless channel to estimate distances.

Given the problem above, a classical/intuitive solution, proposed by us, would be the target receiving a packet at each time with its correspondent RSSI value to establish how close it is from a sink. In a similar way, our novel proposal also uses the RSSI attribute, but more times than the classical solution, employing an algorithm which use the Kalman filter basis.

3 Proposed solution

Taking into account the figure 1, our proposed solution would estimate when the target is close to a sink based on d_1, d_2 and D distances, the angular velocity W_k and the angle variation θ_k. Once these variables are calculated, it is possible to estimate when the target will be as close as possible to a sink.

The d_1 and d_2 distances can be inferred through the RSSI measured when the sinks and the target interchange control messages. However, a RSSI value can be unstable since it is affected by the wireless channel, in other words, by the noise introduced by the channel itself. In this sense, it could be not enough to estimate the distance based only on a unique RSSI value, because it could be a wrong estimation. For this reason, in order to do a suitable estimation of distances, it is required to collect more than one RSSI values, which, in turn, will provide proper estimated values of w_k and θ_k variables. The quality of these variables is very important since it will allow to estimate properly the moment in which the target should send its data to a sink. As a result,

the more RSSI values are sampled, the higher quality of w_k and θ_k is obtained. However, the above will produce more energy consumption because more control packets are transmitted, and then, it has to be considered at setting parameters for the simulation.

In addition, it is necessary to determine how many control packets should be sent in order to obtain suitable distance estimations and, at the same time, saving energy consumption to extend the lifetime of the sensor, and avoiding the noise negative effects.

In terms of energy consumption, it is necessary to consider consumption by transmission and reception in the sensor. For transmission, the energy consumption is defined as $(E_{elec} + E_{amp}) * k * D^n$, while for reception, it corresponds to $E_{elec} * k$. E_{elec} is the energy consumption for codification, modulation and filtering, and E_{amp} corresponds to energy consumption for the $TransmitterPowerAmplifier$ [5]. Additionally, the following parameters values are assumed: E_{elec} is 50nJ/bit, E_{amp} is 100pJ/bit/m^2, n is assumed to be 2 and D corresponds to the distance to send a k-bit packet [6].

In addition to determine the number of packets for distance estimations, it is necessary to define a Δt at which the samples will be taken. If Δt is too high, that is, the frequency to take samples is too low, is highly probable that the closest distance to a sink generated by the solution will be too high respect to the theoretical best distance to a sink. On the contrary, if Δt is too low, the solution will be quite better than the previous, but at the expense of a lot of energy consumption.

3.1 Kalman filter applied to our proposal

The Kalman filter is a recursive algorithm, which is compound of two phases, as shown in the figure 2, that are performed iteratively: a Prediction phase and a Correction phase. For space reasons, the Kalman filter basis are omitted, but if you are interested in the basis details, you can refer to [7] [8].

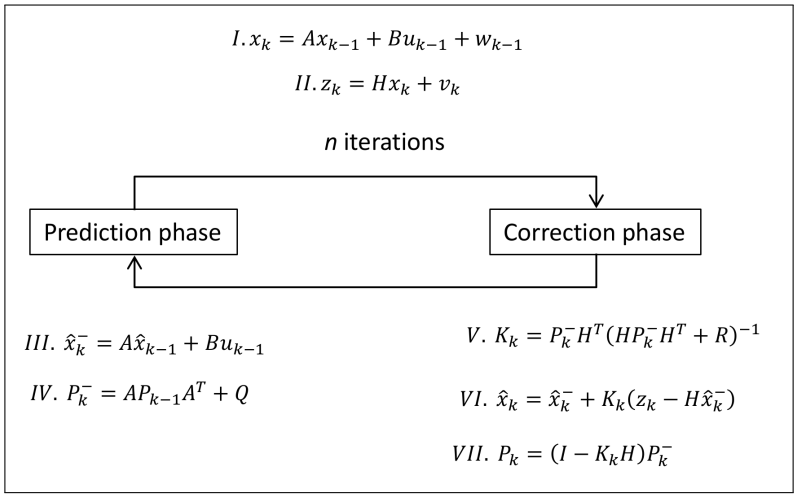


Figure 2: Kalman filter algorithm phases.

We propose to perform two Kalman Filters as it is shown in the Figure 3. One Kalman filter for RSSI and distances (d_1 and d_2), and the other one for θ_k and w_k estimations. For the first one, since we are assuming the RSSI and distances estimated should converge to a constant value, the matrix A, indicated in Figure 2 expressions (I), (III) and (IV), must be equal to "1" according to the Kalman Filter theoretical basis. In addition, the observation matrix H also has

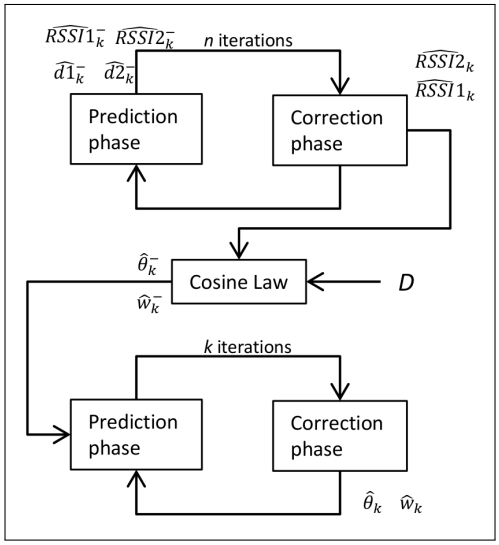


Figure 3: Kalman filter block diagram implementation.

a "1" value, because only one variable is observed. For the second Kalman filter, the expression (I) in the Figure 2 can be defined as follows, where A matrix can be inferred:

$$x_k = \begin{bmatrix} \theta_k \\ w_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_{k-1} \\ w_{k-1} \end{bmatrix} + w_{k-1} \tag{1}$$

In the expression above, notice that the angular velocity is indicated in uppercase and the process noise in lowercase. The A matrix expressed in (1) indicates the transition model state which, basically, defines the circular trajectory of the target. Notice that Δt indicates the time in which we want generate the next k^{th} state of the system. The Δt should not be selected arbitrary since a too high value of Δt will generate excellent results but at the expense of a high energy consumption, and a too low value of Δt will not afford accurate target position.

In the $Algorithm1$ is presented the main aspects of our solution. In addition, a block diagram in the figure 3 is shown to complement the pseudocode explanation.

In line 2, the target is moved with an angular velocity w_s, an angle displacement θ_s during a lapse of Δt_s for the movement simulation of the sensor, which is indicated with the index s. In line 3, Δt_k indicates the lapse in which the samples for Kalman algorithm will be taken. In 6, Paq_{tx} indicates the number of packets with RSSI values in order to do a proper estimation using a Kalman algorithm, that is, to generate the values indicated in line 8 and 9. In 10, it is predicted the RSSI for s_1 and s_2 according to the previous corrected values using a Kalman filter. In 11, the estimated distances to s_1 and s_2 are predicted based on the estimated RSSI previously obtained.

In 13, a Kalman filter is applied over $\hat{d1}_k^-$ and $\hat{d2}_k^-$ to obtain the corrected values $\hat{d1}_k$ and $\hat{d2}_k$. Once the distances has been calculated in 13, in line 14 the cosine law is applied to calculate $\bar{\theta_k}$ based on the values of $\hat{d1}_k$, $\hat{d2}_k$, r and D (Figure 1). Due to $\bar{\theta_k}$ and Δt_k is possible to obtain $\bar{W}_k$ based on the expression (1). Notice that $\bar{\theta_k}$ and Δt_k correspond to the observed values z_k indicated in the expression (II) of the Figure 2. In line 15 and 16, based mainly on observed values $\bar{\theta_k}$ and $\bar{W}_k$; the transition model state matrix A described in the expression (1); and through the Kalman filter theory is possible to predict and correct the values of θ and W, that is, $\hat{\theta}_k^-$ and $\hat{W}_k^-$, and $\hat{\theta}_k$ and $\hat{W}_k$ respectively. Once these values are calculated the (x, y) coordinates can be found and determine how close is the target from the sinks.

Algorithm 1 Solution Pseudocode

while $Bat_{sensor} > 0$ **do**
  Move w_s, θ_s
  Set Δt_k
  if Δt_k *is triggered* **then**
    Calculate d_1 *and* d_2
    Set $n = Paq_{tx}$
    for $i = 1$ *to* n **do**
      Generate $RSSI1$, $RSSI2$ + *White Noise*
      Send RSSI packets to the sinks
      Predict $\widehat{RSSI1_k^-}$, $\widehat{RSSI2_k^-}$
      Predict $\widehat{d1_k^-}$, $\widehat{d2_k^-}$
    end for
    Correct $\widehat{d1_k}$, $\widehat{d2_k}$
    Calculate $\overline{\theta_k}$, $\overline{w_k}$
    Predict $\widehat{\theta_k^-}$, $\widehat{w_k^-}$
    Predict $\widehat{\theta_k}$, $\widehat{w_k}$
  end if
end while

4 Results

Before to show the results, it is necessary to present the following figures, which represent some performance aspects of the traditional and our proposed solution.

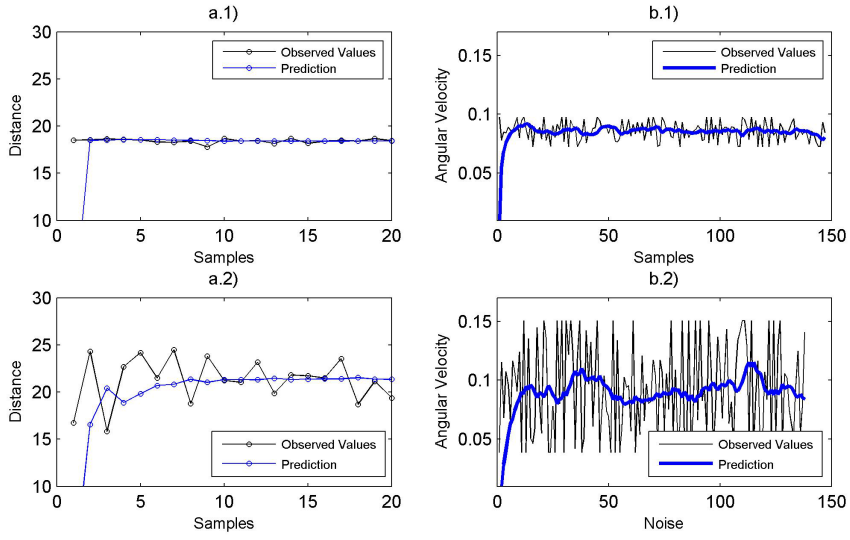


Figure 4: Performance results for traditional and our proposed solution.

The figures a.1) and a.2) show the prediction based on Kalman for a noise of 0.01 and 0.1, respectively, and for 20 packets to determine a distance at a given time. In the figure a.1), as the noise is lower than a.2), the prediction is more accurate than a.2). In the figures b.1) and b.2) is presented the delta angular velocity for a noise of 0.01 and 0.1, respectively, in order to determine the target position in a given time. In b.2), as the noise is quite high, the prediction is more variable than b.1). As a result, in b.1) the target position is more accurate than b.2).

For c) and d) figures, the main simulation parameters were: $\Delta t_k = 700ms$, $Paq_{tx} = 200$, $DataSizeforTransmission$ equal to $500KB$, $Perimeter = 400m$ and $TargetSpeed = 5.52m/s$. In c) it is presented the energy consumption for both traditional and our proposed solution according to different noise values. For 0.1, 0.2, and 0.3 noise values, the traditional has a lesser energy consumption than our proposed solution. However, for values higher than 0.3, our proposed solution presents a lesser energy consumption because the traditional solution behaviour is affected by the high noise level. The performance is affected due to sending extra data transmission caused by the high noise present in the communication channel. In d), the distance from the target and the sink 2 is calculated at the moment of data transmission for both traditional and our proposed solution. From 0.1 to 0.3 the performance for both are quite similar. However, for values higher than 0.3 our algorithm outperforms classic solution since it less prone to disturbances caused by noisy wireless channels.

5 Conclusions

In this paper, we propose a predictive algorithm based on Kalman filter techniques to estimate the proper time at which the sensor is close as much as possible to a sink, in order to reduce the energy consumption in the sensor.

We propose an energy efficient routing algorithm based on Kalman Filtering techniques to predict a future state of the system in order to determine the accurate time to send data information from a sensor to a sink. The theoretical basis and operation of our algorithm is explained in detail, and its performance is evaluated for an athlete monitoring scenario for illustrative reasons.

The main conclusion of our proposed solution is about its performance in noisy communication channels. It means that the energy consumption of our algorithm is lesser than the traditional method, presenting less impact for noisy communication channels.

Acknowledgement

This research paper was made possible through the help and support from the European Union Project GOLDFISH at Framework Programme 7.

Bibliography

[1] I. F. Akyildiz and M. C. Vuran (2010), *Wireless Sensor Networks*, John Wiley & Sons.

[2] J. Zheng and A. Jamalipour (2009); *Wireless Sensor Networks: A Networking Perspective*, Wiley-IEEE Press, ISBN: 978-0-470-16763-2.

[3] D. Gavalas, C. Konstantopoulos, B. Mamalis and G. Pantziou (2010); Mobility Prediction in Mobile Ad Hoc Networks, Book chapter for *Next Generation Mobile Networks and Ubiquitous Computing* by S. Pierre, IGI Global, 226-240.

[4] B. Buchli, F. Sutton and J. Beutel (2012); GPS-Equipped Wireless Sensor Network Node for High-Accuracy Positioning Applications, *Wireless Sensor Networks Lecture Notes in Computer Science*, Springer, 7158: 179-195.

[5] S. Li, X. Ma, X. Wang, M. Tan (2011); Energy-efficient multipath routing in wireless sensor network considering wireless interference, *Journal of Control Theory and Applications*, 9(1): 127-132.

[6] P. S. Boluk, S. Baydere, A. E. Harmanci (2011); Robust Image Transmission Over Wireless Sensor Networks, *Mobile Networks and Applications*, 16(2): 149-170.

[7] P. S. Maybeck (1979); Stochastic Models, Estimation and Control, *Mathematics in Science and Engineering*, Vol.141.

[8] G. Welch and G. Bishop (1995); An Introduction to the Kalman Filter. Technical Report. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):508-519, August, 2015.

A Visualization Technique for Accessing Solution Pool in Interactive Methods of Multiobjective Optimization

E. Filatovas, D. Podkopaev, O. Kurasova

Ernestas Filatovas*, Olga Kurasova
Vilnius University
Institute of Mathematics and Informatics,
Akademijos str. 4, LT–08663 Vilnius, Lithuania
Ernest.Filatov@gmail.com, Olga.Kurasova@mii.vu.lt
*Corresponding author: Ernest.Filatov@gmail.com

Dmitry Podkopaev
1. Systems Research Institute
Polish Academy of Sciences
Newelska str. 6, 01-447 Warsaw, Poland
2. University of Jyväskylä
Departament of Mathematical Information Technology
P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland
Dmitry.Podkopaev@gmail.com

Abstract: Interactive methods of multiobjective optimization repetitively derive Pareto optimal solutions based on decision maker's preference information and present the obtained solutions for his/her consideration. Some interactive methods save the obtained solutions into a solution pool and, at each iteration, allow the decision maker considering any of solutions obtained earlier. This feature contributes to the flexibility of exploring the Pareto optimal set and learning about the optimization problem. However, in the case of many objective functions, the accumulation of derived solutions makes accessing the solution pool cognitively difficult for the decision maker. We propose to enhance interactive methods with visualization of the set of solution outcomes using dimensionality reduction and interactive mechanisms for exploration of the solution pool. We describe a proposed visualization technique and demonstrate its usage with an example problem solved using the interactive method NIMBUS.
Keywords: Multiobjective optimization, interactive methods, Pareto front visualization, dimensionality reduction, multidimensional scaling.

1 Introduction

Many real-world optimization problems are multiobjective in their nature. Existence of a solution optimizing all objectives simultaneously is unlikely in practice. Solving a multiobjective problem is usually understood as finding a Pareto optimal solution which is the most preferred for a decision maker (DM) [16]. Therefore most of the multiobjective optimization methods rely on preference information obtained from the DM. We concentrate on *interactive methods* consisting of iterations, where at each iteration the DM provides some preference information and receives solution(s) derived based on this information as feedback [3].

In many interactive methods, a solution obtained at each iteration serves as the starting point for the next iteration. The DM considers this solution and expresses preference information aiming at its improvement. It is useful to allow the DM selecting for consideration the solution obtained not only during the last iteration but any previous iteration. Few interactive methods include this possibility (see e.g. [8, 22]) and to our knowledge, none of the papers describe the selection process in detail. The issue to be addressed here is high cognitive load placed on the

DM when dealing with many-objective problem and a large number of accumulated solutions. Even though the number of iterations in an interactive solution process can be moderate [7], many interactive methods such as Tchebycheff method [33], NIMBUS [22] and others generate several Pareto optimal solutions at each iteration, which makes the solution set grow fast.

In this paper, we address the problem of assisting the DM in selecting one among many Pareto optimal solutions at each iteration of an interactive method in the case of a large number of objectives. We propose to enhance interactive methods with a technique which visualizes a set of the obtained solution outcomes and provides interactive mechanisms of its exploration. This technique is based on multidimensional scaling method [2] which maps the solution outcomes on a plane while trying to preserve distances between them. Despite significant lost of information, such dimensionality reduction creates a holistic view on the outcome set helping the DM to develop a cognitive map. This method has been successfully applied earlier for visualizing solution outcomes in multiobjective evolutionary algorithms [10], but never in interactive methods. Its integration into interactive methods requires some modification for handling inclusion of new solutions into the solution pool.

This paper is organized as follows. In Section 2 we present background information and studies underlying our research, namely basic notions of multiobjective optimization, a general structure of interactive methods and the employed approach to dimensionality reduction. In Section 3 we propose a visualization technique which can be integrated in many interactive methods. In order to demonstrate this technique in action, in Section 4 we integrate it to the interactive method NIMBUS, and use the technique for solving an example problem. Section 5 presents conclusions.

2 Background

2.1 The problem of multiobjective optimization

The multiobjective optimization problem is formulated as follows [16]:

$$\min_{\mathbf{x}\in S} \mathbf{f}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is a *decision (variable) vector*; $S \subseteq \mathbb{R}^n$ is a set of *feasible solutions*; $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x}))^{\mathrm{T}}$, $f_i : \mathbb{R}^n \to \mathbb{R},\ i \in \{1, 2, \ldots, k\}$ are *objective functions*, $k \geq 2$. For each decision vector $\mathbf{x} \in S$, $\mathbf{f}(\mathbf{x})$ is called its *outcome*.

A decision vector $\mathbf{x}^*$ is called *Pareto optimal solution* of problem (1) (or *Pareto optimum* for short), if there does not exist another vector $\mathbf{x} \in S$ such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$ for all $i \in \{1, 2, \ldots, k\}$ and $f_j(\mathbf{x}) < f_j(\mathbf{x}^*)$ for at least one j. The set of Pareto optimal solutions is called *Pareto optimal set*, and the set of outcomes of all Pareto optimal solutions is called *Pareto front*.

Solving the problem (1) is commonly understood as finding a Pareto optimum $\mathbf{x}^*$ whose outcome $\mathbf{f}(\mathbf{x}^*)$ is the most preferred for the DM (see e.g. [3, 16]). Therefore, most of multiobjective optimization methods use preference information provided by the DM. In this paper, we concentrate on *interactive methods* for they are believed to be most promising methods of multiobjective optimization because of numerous advantages [25]. An important advantage of interactive methods is the possibility for the DM to learn about the problem during the solution process, which makes him/her more confident in the final choice [1].

2.2 Interactive methods

In interactive methods [16, 18], the DM provides preference information progressively during the solution process and obtains Pareto optima derived based on this information as feedback.

Many interactive methods have the underlying idea of exploring the Pareto optimal set by shifting DM's attention from one solution to another. The notion of *current solution* is introduced as the Pareto optimum considered by the DM in the current iteration. The DM expresses preferences aiming (explicitly or implicitly) at improving the current solution, i.e. finding a more preferred Pareto optimum. A Pareto optimum derived based on these preferences can be considered as the current solution in the next iteration. This process of producing a sequence of Pareto optima can be viewed as moving from one Pareto optimum to another, where preference information determines the direction of movement.

In the case of a cognitively complex model underlying a multiobjective optimization problem, the DM cannot be sure that his/her preference information always leads to a more preferred solution. Besides that, sequential improvement of Pareto optimal solutions based on DM's preferences cannot guarantee the convergence to the most preferred solution among the whole Pareto optimal set, by analogy with global optimization problems where a series of local improvements of a feasible solution does not necessary lead to a global optimum. Therefore it is reasonable to give the DM possibility of exploring different parts of the Pareto optimal set by going back to solutions derived earlier and starting moving from them in different directions. This possibility can be implemented by saving all the obtained Pareto optima in the *solution pool* and enabling the DM in each iteration selecting any element of the pool as the current solution for the next iteration. Thus instead of a sequence of Pareto optima one can have a tree where each branch relates to such a move back.

A general structure of interactive methods including the possibility of moving back is described below step-by-step (Algorithm 1).

Algorithm 1

Step 0. **Initialize**. Set initial variable values and derive Pareto optimal solution(s) using an initial preference model (based on some initial DM's preference information or a rule of thumb).

Step 1. **Update**. Add the obtained solution(s) into the pool.

Step 2. **Select**. Ask the DM to select a solution from the pool.

Step 3. **Preference expression**. If the DM regards the selected solution the most preferred, then the method stops with the selected solution as the final one. Otherwise set the selected solution as the current solution of the next iteration and ask the DM to express preferences aiming at its improvement.

Step 4. **Solution derivation**. Build the preference model based on DM's preference information and derive new Pareto optimal solution(s) using this model. Go to Step 1.

Some authors mention the possibility of selecting the current solution from the solution pool in interactive methods (see e.g. [6, 8, 22, 30]). Many interactive methods can be easily enhanced by adding such possibility and thereby be fit into the structure described above. However to our knowledge, no papers present any details about implementing the procedure of solution selection.

As the size of the solution pool grows, it becomes cognitively more difficult for the DM to keep and operate with it in his/her mind. Indeed, comparing and choosing among a number of objects exceeding the mental capacities obviously leads to mistakes and inconsistencies as illustrated e.g. in [32]. Therefore the DM can benefit from a graphical tool visualizing the solution pool and assisting with solution selection.

2.3 Graphical representation of solution outcomes

It must be noted that most of the multiobjective optimization methods rely on the assumption that the DM compares different solutions and chooses the most preferred one only based on their outcomes. Therefore most of existing visualization techniques (see e.g. [10, 12, 13, 17]) concentrate on presenting the set of the Pareto optimal outcomes in the so-called *objective space* $\mathbb{R}^k$. Visualizing a set of Pareto optimal outcomes is a non-trivial task in the case $k > 3$. If the solution set is relatively small, then all components of all outcomes can be presented in a diagram such as scatter plot, bar charts, spider-web charts and others (see comprehensive surveys in [9,17]). For bigger solution sets, such representations enforce the DM to deal with abundance of information. Then dimensionality reduction methods can be helpful.

As dimensionality reduction involves loss of information, the problem of extracting information which remains can be approached in different ways depending on the aim the visualization serves. For example, in self-organizing maps [36], artificial neural networks are trained to preserve the topological properties of the input space. Principal component analysis (PCA) [15] tries to preserve variances of data. Interactive decision maps [4, 12] are based on two-dimensional projections, they are also used for the visualization of Pareto outcomes.

Multidimensional scaling method (MDS) [2, 5] aims at finding points $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ in the low-dimensional space $\mathbb{R}^s$, $s < k$, such that the distances between them are as close as possible to the distances between the original points $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m$ in the multidimensional space $\mathbb{R}^k$. This is achieved by minimizing the stress function

$$E_{MDS} = \sum_{i<j} \Big(d(\mathbf{v}_i, \mathbf{v}_j) - d(\mathbf{z}_i, \mathbf{z}_j) \Big)^2. \tag{2}$$

Here $d(\cdot, \cdot)$ is the distance between two points in the corresponding space. The procedure of finding points $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ by minimizing (2) is called *mapping*. We propose to use MDS for visualizing the solution pool in interactive methods because it detects the similarity between Pareto optimal outcomes, and allows the DM distinguishing areas by how much they have been explored. This helps in focusing the DM's attention on solutions located in preferred parts of the Pareto front.

When using MDS during an interactive solution process, a problem of mapping new solution outcomes occurs. After adding them to the solution pool and recalculating low-dimensional points by minimizing (2), the points corresponding to already mapped solution outcomes will also change. In order to avoid this changing, so-called *relative mapping* [27] can be applied, where only coordinates of the newly added points are calculated. The stress function of relative mapping is following:

$$E_R = \sum_{i,j=m+1, i<j}^{m+\hat{m}} \Big(d(\hat{\mathbf{v}}_i, \hat{\mathbf{v}}_j) - d(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j) \Big)^2 + \sum_{i=m+1}^{m+\hat{m}} \sum_{j=1}^{m} \Big(d(\hat{\mathbf{v}}_i, \mathbf{v}_j) - d(\hat{\mathbf{z}}_i, \mathbf{z}_j) \Big)^2, \tag{3}$$

where $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m$ are the previously mapped multidimensional points and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are their corresponding lower-dimensional points which are fixed, $\hat{m}$ is the number of new points, $\hat{\mathbf{z}}_{m+1}, \hat{\mathbf{z}}_{m+2}, \ldots, \hat{\mathbf{z}}_{m+\hat{m}}$ and $\hat{\mathbf{v}}_{m+1}, \hat{\mathbf{v}}_{m+2}, \ldots, \hat{\mathbf{v}}_{m+\hat{m}}$ are the new points and their corresponding lower-dimensional points, respectively.

However, after applying the relative mapping, the value of the stress function E_{MDS} calculated for all the solution outcomes by formula (2) increases, which indicates deterioration of mapping quality. Therefore, after visualizing a large number of the new outcomes using relative mapping, all the mapped points must be recalculated by the MDS method again, i.e. *remapping* must be performed.

Our previous research was devoted to usage MDS for visualizing the Pareto front in multiobjective evolutionary algorithms [10]. A similar approach was developed in [34] for reviewing the set of obtained Pareto optimal solutions when solving the problem (1). However, the possibility of integrating MDS into interactive methods has not been studied before. In the next section, we describe the proposed visualization technique based on MDS and its integration into interactive methods.

3 Enhancement of interactive methods with visualization

The general structure of interactive methods described in Subsection 2.2 is modified and presented as a scheme in Figure 1. The visualization tool is used in Step 2 for assisting the DM in exploring the solution pool and selecting from it a Pareto optimal solution to be considered as the most preferred solution or the current solution in the next iteration. In order to enable using the visualization, a visual representation of the solution pool is constructed or updated in Step 1. We propose to allow the DM choosing, whether to use the visualization technique or select the solution in a traditional way implemented in the method. Thus, adding the visualization in the method does not imply any additional efforts from the DM. If the DM finds the visualization useless, he/she can ignore it which means that its integration into the method does not deteriorate DM's experience.

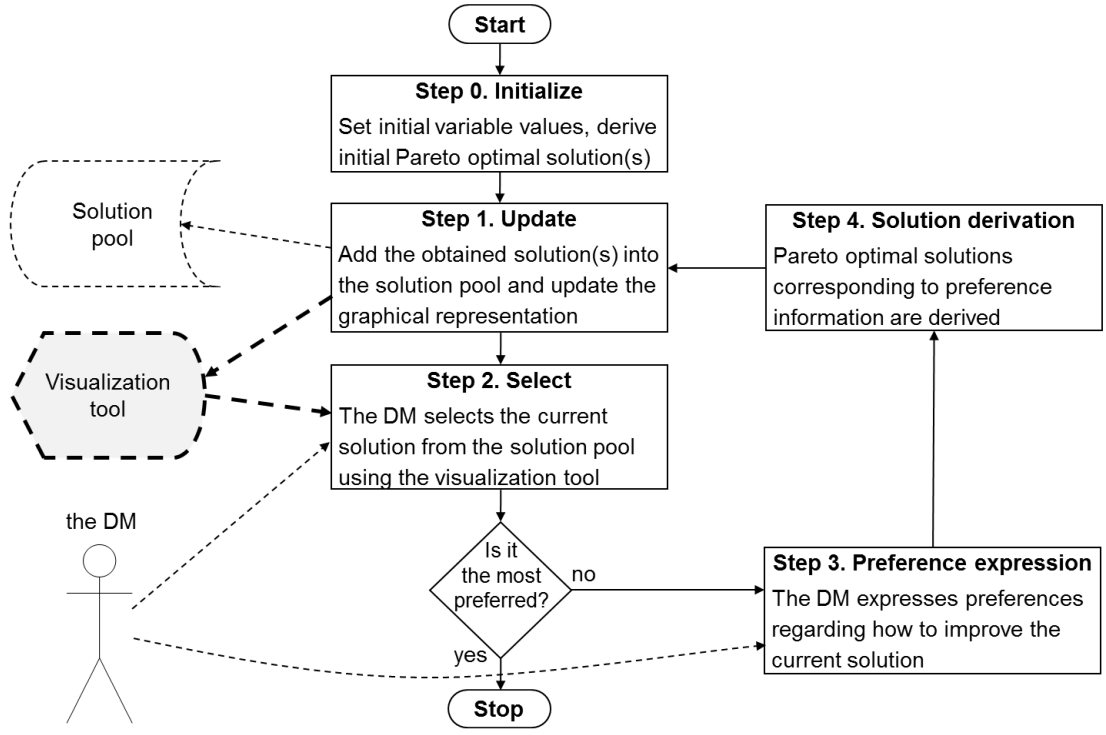


Figure 1: A general enhanced scheme of interactive methods.

The proposed visualization technique consists of mapping of solution outcomes on a plane using the MDS method and interacting with the DM for exploring the outcome set and selecting a solution. Our approach does not depend on the interactive method used, therefore the technique can be implemented for a wide range of interactive methods. We propose to visualize outcomes from the solution pool as a scatter plot of two-dimensional points mapped using the MDS method or relative mapping, that can be implemented in visualization tool (see Figure 1). The following

two operations will be used for constructing and updating this visualization:

- **map all solution outcomes** – given outcomes $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m$ of all Pareto optimal solutions from the solution pool, corresponding two-dimensional points $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are calculated by minimizing the MDS stress function (2) as described in Subsection 2.3. Here m is the number of solution outcomes, $\mathbf{z}_j \in \mathbb{R}^k$ and $\mathbf{v}_j \in \mathbb{R}^2$ for all $j = 1, \ldots, m$.
- **map new solution outcomes** – given outcomes $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m$ of Pareto optimal solutions from the solution pool mapped earlier, their corresponding two-dimensional points $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$, and the outcomes of newly derived Pareto optimal solutions $\hat{\mathbf{z}}_{m+1}, \hat{\mathbf{z}}_{m+2}, \ldots, \hat{\mathbf{z}}_{m+\hat{m}}$, the two-dimensional points $\hat{\mathbf{v}}_{m+1}, \hat{\mathbf{v}}_{m+2}, \ldots, \hat{\mathbf{v}}_{m+\hat{m}}$ corresponding to the new solutions are calculated by minimizing the relative mapping stress function (3) as described in Subsection 2.3.

In order to evaluate similarities between the obtained solution outcomes the MDS method requires several points. Therefore the visualization starts being constructed after a predefined number of solutions m has been collected in the solution pool. In the case of a small number of solutions, it can be easier for the DM to select one of them by straightforward comparison of the solution outcomes rather than studying the two-dimensional visualization. Therefore we propose to set initial value of m not smaller than 7.

Once new Pareto optimal solutions are derived in Step 4 (Figure 1) and added to the solution pool, the visual representation is updated in Step 1 either by mapping all solutions or only new solutions. In the first case, the locations of the previously mapped solutions will also change which may distort the cognitive map of the solution pool the DM has developed. On the other hand, mapping only new solutions may affect the accuracy of correspondence between distances in the objective space and two-dimensional space. Therefore, we propose to use a compromise strategy where new solutions are mapped during most of iterations, but the map of all solutions is recalculated on DM's demand or after accumulation of a large number of new solutions. When the number of the added solutions l reaches the predefined threshold limit m_p, the map of all Pareto optimal outcomes is recalculated by applying the MDS method again. We propose to set the limit $m_p = 0.5m$, where m is the number of outcomes already mapped by MDS. Hence, the higher value m is, the less frequent remapping by MDS is required. The detailed procedure of constructing the map of the obtained outcomes is presented as Algorithm 2.

Algorithm 2 Constructing the map of the solution outcomes

Input: The solution outcomes $\mathbf{z}_1, \ldots, \mathbf{z}_m$, obtained by using an interactive method, where m is the number of outcomes.

Output: The 2D map (scatter plot).

Step 1: Set the number of solution outcomes mapped by relative MDS to $l = 0$.

Step 2: The dimensionality of the solution outcomes $\mathbf{z}_1, \ldots, \mathbf{z}_m$ is reduced to 2 by the MDS method, and the two-dimensional points $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are displayed in a 2D scatter plot and presented to the DM. Set the value of threshold limit $m_p = 0.5m$.

Step 3: When $\hat{m}$ new solution outcomes $\hat{\mathbf{z}}_{m+1}, \ldots, \hat{\mathbf{z}}_{m+\hat{m}}$ are obtained by the interactive method, two-dimensional points $\hat{\mathbf{v}}_{m+1}, \ldots, \hat{\mathbf{v}}_{m+\hat{m}}$ are calculated by the relative mapping, they are added to the 2D scatter plot and we set $m = m + \hat{m}$, $l = l + \hat{m}$. If the DM is satisfied with the obtained solution, then go to Step 5.

Step 4: Check if $l > m_p$ or the DM wants to remap the solution outcomes. If Yes then go to Step 1, else go to Step 3.

Step 5: Stop.

The visualization should be accompanied with graphical tools and interactive mechanisms based on graphical user interface, allowing the DM:

- to see additional information related to the way solution were obtained, e.g. information about the extent at which solutions satisfy preferences based on which they were derived, relations between solutions related to the sequence in which they were obtained;
- to perform operations on visualized outcomes, e.g. hide selected outcomes or outcomes satisfying a given property in order to focus on the most interesting outcomes, recalculate mapping on demand;
- to get full information about the solution corresponding to any selected outcome.

The implementation of the above enhancements of the visualization might depend on the interactive method used as well as the graphical user interface of the method implementation.

4 Example implementation

In this section, we demonstrate the proposed visualization technique by solving an example problem using a popular interactive multiobjective optimization method NIMBUS [16, 19]. It is implemented in the IND-NIMBUS software platform ([28, 29]) designed for solving industrial problems, as well as in a web-based application WWW-NIMBUS available via the Internet [21]. The NIMBUS method has been applied to a number of engineering problems [11, 22, 24].

In the NIMBUS method, new Pareto optimal solutions are derived by solving scalarized problems based on preference information provided by the DM in the form of a classification. In each iteration of this method, the DM considers the current Pareto optimal solution $\mathbf{x}^{\mathrm{c}} \in S$ and corresponding outcome and expresses preferences by classifying the objective functions $f_i(\mathbf{x}^{\mathrm{c}})$, $i = 1, 2, \ldots, k$ into up to the five classes related to five types of desirable or acceptable changes of the function values [22]: $I^{<}$ – to be improved (i.e. decreased) as much as possible; $I^{\leq}$ – to be improved till some desirable aspiration level $\bar{z}_i < f_i(\mathbf{x}^{\mathrm{c}})$; $I^{=}$ – the current function values are acceptable; $I^{\geq}$ – may be impaired (i.e. increased) till some upper bound $a_i > f_i(\mathbf{x}^{\mathrm{c}})$; $I^{\diamond}$ – the function values are allowed to change freely. This classification must be done in such a way that at least one objective function value should be improved, and at least one is allowed to impair. The classification is used to build scalarizing functions for deriving new Pareto optimal solutions. In synchronous NIMBUS [22], four different scalarization functions are used in order to present to the DM diverse solutions related to expressed preferences, thereby up to four different solutions can be obtained according the same preference information.

4.1 Numerical example

As an example for demonstrating the proposed visualization technique in use, we chose the five-objective river pollution problem which was implemented in IND-NIMBUS. The problem was originally presented in [26] and modified in [20], where the fifth (nonsmooth) objective function was included. The mathematical formulation of the problem is as follows:

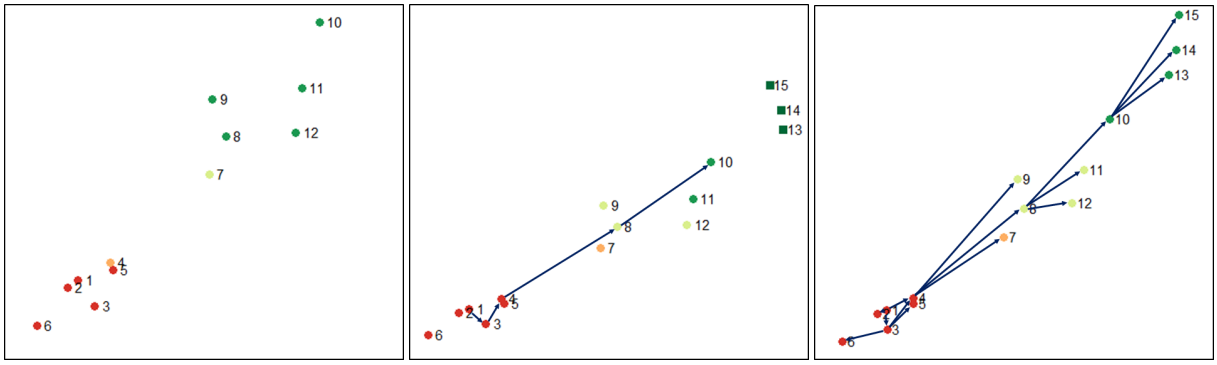
$$
\begin{aligned}
&\text{maximize } f_1(\mathbf{x}) = 4.07 + 2.27x_1 \\
&\text{maximize } f_2(\mathbf{x}) = 2.60 + 0.03x_1 + 0.02x_2 + \frac{0.01}{1.39 - x_1^2} + \frac{0.30}{1.39 - x_2^2} \\
&\text{maximize } f_3(\mathbf{x}) = 8.21 - \frac{0.71}{1.09 - x_1^2} \\
&\text{minimize } f_4(\mathbf{x}) = -0.96 + \frac{0.96}{1.09 - x_2^2} \\
&\text{minimize } f_5(\mathbf{x}) = \max[|x_1 - 0.65|, |x_2 - 0.65|] \\
&\text{subject to } 0.3 \leq x_1, x_2 \leq 1.0.
\end{aligned} \tag{4}
$$

It describes a (hypothetical) pollution problem of a river, where a fishery and a city are polluting water. The decision variables x_1 and x_2 represent the proportional amounts of biochemical oxygen demanding material removed from water in two treatment plants located after the fishery and after the city [23]. Here f_1 and f_2 describe the quality of water after the fishery and after the city, respectively; f_3 and f_4 show the percent return on investment at the fishery and the addition to the tax rate in the city, respectively; and f_5 describes the functionality of the treatment plants. Hence, the first three objective functions must be maximized, while the fourth and fifth objective functions are to be minimized.

The DM's strategy and solution process when solving the original four-objective river pollution problem [26] using NIMBUS method was described in [14]. Below we describe the process of solving the problem (4) using synchronous NIMBUS implemented in IND-NIMBUS, where the DM gets additional support from our visualization technique. When addressing multiobjective optimization problems the DM can have different priorities across objectives in mind. Further we assume that the DM has higher priority to improving the quality of the water after the fishery and the city. The solution process described below is divided into several stages according to the DM's actions related to the usage of the visualization technique:

Stage 1. When starting solving the problem by using the IND-NIMBUS, at first a so-called neutral compromise solution is obtained using an initial preference model and presented to the DM. Naturally the DM is not satisfied with the obtained solution, therefore he/she tries to improve it by making several different classifications of the objective functions. During the solution process the DM selects the most preferred solution from the obtained four solutions as the current solution and tries to improve it in the next iteration. This process is repetitive, and the solution pool is constantly updated with the new derived solutions.

Stage 2. After a predefined number of the solutions is collected in the solution pool, the visualization technique is started to be used by the DM (see Figure 2a). Here the points of the map correspond to the obtained solution outcomes. We omit drawing axes of the two-dimensional coordinate system due to the fact that the coordinates do not have individual meaning, as the only important aspect of mapping is preserving distance between points. The points of the scatter plot are coloured using RdYlGn colour scheme, where colour of each mapped point represents its distance to the reference point that express the current DM's preferences (red colour corresponds to large distance, green – to small distance). The numbers near the points mean the order in which the corresponding solutions were obtained. The presented map of obtained outcomes is constructed by using the MDS method, when the 8th solution has been selected as the current one. Now the DM can explore the map, see the relations between the outcomes and think about the further directions of the solution process. It can be noticed from the map that the 10th solution is located furthest from the area of the objective space where the DM started the solution process. As the DM wants to explore the objective space in direction further away from first solution, he decides to select the 10th solution as the current one.



(a) (b) (c)

Figure 2: Two-dimensional map of the Pareto optimal solution outcomes, constructed by the visualization technique.

Stage 3. The DM tries to improve the 10th solution by using a different classification of objective functions. Then the solution pool is updated by new outcomes and they are added to the map by using the relative mapping (see Figure 2b). The points mapped using MDS have circular shape while the new points are presented as squares. Here the graph of the solution process is also displayed (the arrows are drawn between the points corresponding to obtained outcomes which once served as current solutions). The location of new points confirms that the selection of 10th solution as the current solution and provision of the preference information allowed the DM to advance in exploring the objective space, e.g. obtain Pareto optimal solutions which demonstrate bigger differences from earlier obtained solutions. Afterwards, the DM wants to remap all the obtained outcomes by MDS in order to see relations between them more precisely, and the updated map is presented (see Figure 2c). After remapping, the places of the points on the map have been changed slightly comparing with the previous ones, which does not affect the cognitive map developed in the DM's mind. However, after remapping the furthest location of 15th solution outcome from earlier solutions is more expressed. Here the arrows are drawn between all the solutions from the solution pool in sequence order in which the solutions were obtained (see Figure 2c). Next, the DM can select other solution from the solution pool as the current solution for further improvement, change the solution strategy by giving higher priority to other objectives, and search for more preferable solutions in unexplored parts of the decision space with the support of the MDS-based visualization technique, or using only the original IND-NIMBUS user interface. Hence, the obtained map helps the DM to keep in mind a holistic image of the current state of the solution pool, as well as the history of the solution process.

It must be pointed out that the purpose of this example is not to describe the whole solution process of the problem, but to demonstrate the performance and flexibility of the proposed visualization technique. Hence, the three presented stages cover the main aspects of its usage during the solution process.

5 Conclusions

The core idea of interactive methods is using a Pareto optimum derived during each iteration (so-called current solution) as the starting point for the next iteration. The possibility of selecting Pareto optima derived in any of previous iterations as the current solution contributes to the flexibility and freedom of Pareto optimal set exploration. This aspect of implementing interactive methods has not received any attention in the literature.

In this paper we have proposed to enhance interactive methods with a technique visualizing the outcomes from the solution pool and assisting the DM in its exploration and solution selection. The resulting visualization is presented in the form of an interactive 2D scatter plot obtained using MDS method, which helps the DM to detect similarity between outcomes and focus on more preferred parts of the Pareto front. This visualization facilitates the DM in developing a cognitive map of the set of solution outcomes. We have described an general approach of creating and integrating the visualization into interactive methods, and also illustrated it with a simple example. However the visualization should be implemented using a graphical user interface and, depending on the method, enhanced with graphical and interactive elements improving DM's experience.

Acknowledgment

Ernestas Filatovas is supported by the postdoctoral fellowship funded by European Union Structural Funds project "Postdoctoral Fellowship Implementation in Lithuania".

Bibliography

[1] Belton, V., Branke, J., Eskelinen, P., Greco, S., Molina, J., Ruiz, F., Słowiński, R. (2008); Interactive multiobjective optimization from a learning perspective, in Branke, J., Deb, K., Miettinen, K., Słowiński, R., editors, *Multiobjective Optimization (Lecture Notes in Computer Science, 5252)*, Springer, ISBN 978-3-540-88907-6, 405–433.

[2] Borg, I., Groenen, P. J. (2005); *Modern multidimensional scaling: Theory and applications*, Springer, ISBN 978-0-387-25150-9.

[3] Branke, J., Deb, K., Miettinen, K., Słowiński, R. (2008); *Multiobjective optimization: Interactive and evolutionary approaches (Lecture Notes in Computer Science 5252)*, Springer, ISBN 978-3-540-88908-3.

[4] Castelletti, A., Lotov, A. V., Soncini-Sessa, R. (2010); Visualization-based multi-objective improvement of environmental decision-making using linearization of response surfaces, *Environmental Modelling & Software*, ISSN 1364-8152, 25(12): 1552–1564.

[5] Dzemyda, G., Kurasova, O., Žilinskas, J. (2013); *Multidimensional Data Visualization: Methods and Applications (Springer Optimization and Its Applications, 75)*, Springer, ISBN 978-1-4419-0235-1.

[6] Filatovas, E., Kurasova, O. (2011); A decision support system for solving multiple criteria optimization problems. *Informatics in Education*, ISSN 1648-5831, 10(2): 213–224.

[7] Gardiner, L., Vanderpooten, D. (1997); Interactive multiple criteria procedures: Some reflections. In Climaco, J., editor, *Multicriteria Analysis*, Springer, ISBN 978-3-642-64500-6, 290–301.

[8] Jaszkiewicz, A., Słowiński, R. (1999). The "light beam search" approach - an overview of methodology and applications. *European Journal of Operational Research*, ISSN 0377-2217, 113(2): 300–314.

[9] Korhonen, P., Wallenius, J. (2008); Visualization in the multiple objective decision-making framework, in Branke, J., Deb, K., Miettinen, K., Słowiński, R., editors, *Multiobjective Optimization (Lecture Notes in Computer Science, 5252)*, Springer, ISBN 978-3-540-88907-6, 195–212.

[10] Kurasova, O., Petkus, T., Filatovas, E. (2013); Visualization of pareto front points when solving multi-objective optimization problems. *Information Technology And Control*, ISSN 1392-124, 42(4): 353–361.

[11] Laukkanen, T., Tveit, T.-M., Ojalehto, V., Miettinen, K., Fogelholm, C.-J. (2010); An interactive multi-objective approach to heat exchanger network synthesis. *Computers & Chemical Engineering*, ISSN 0098-1354, 34(6): 943–952.

[12] Lotov, A., Bushenkov, V., Kamenev, G. (2004); Interactive Decision Maps, *Approximation and Visualization of Pareto Frontier (Applied Optimization, 89)*, Springer, ISBN 978-1-4419-8851-5.

[13] Lotov, A. V., Miettinen, K. (2008); Visualizing the Pareto frontier, in Branke, J., Deb, K., Miettinen, K., Słowiński, R., editors, *Multiobjective Optimization (Lecture Notes in Computer Science, 5252)*, Springer, ISBN 978-3-540-88907-6, 5252: 213–243.

[14] Luque, M., Ruiz, F., Miettinen, K. (2011); Global formulation for interactive multiobjective optimization. *OR Spectrum*, ISSN 0171-6468, 33(1): 27–48.

[15] Mareschal, B., De Smet, Y. (2009); Visual PROMETHEE: Developments of the PROMETHEE & GAIA multicriteria decision aid methods, in *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management, 2009*, e-ISBN 978-1-4244-4870-8, 1646–1649.

[16] Miettinen, K. (1999); *Nonlinear multiobjective optimization (International Series in Operations Research & Management Science, 12)*, Springer, ISBN 978-1-4615-5563-6.

[17] Miettinen, K. (2014); Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectrum*, ISSN 0171-6468, 36(1): 3–37.

[18] Miettinen, K., Hakanen, J., Po (2014a); Interactive nonlinear multiobjective optimization methods. In Ehrgott, M., Figueia, J., and Greco, S., editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, (to appear).

[19] Miettinen, K., Mäkelä, M. (1995); Interactive bundle-based method for nondifferentiable multiobjective optimization: NIMBUS. *Optimization*, ISSN 0233-1934, 34(3): 231–246.

[20] Miettinen, K., Mäkelä, M. M. (1997); Interactive method NIMBUS for nondifferentiable multiobjective optimization problems, in Climaco, J., editor, *Multicriteria Analysis*, Springer, ISBN 978-3-642-64500-6, 310–319.

[21] Miettinen, K., Mäkelä, M. M. (2000); Interactive multiobjective optimization system WWW-NIMBUS on the Internet. *Computers & Operations Research*, ISSN 0305-0548, 27(7): 709–723.

[22] Miettinen, K., Mäkelä, M. M. (2006); Synchronous approach in interactive multiobjective optimization. *European Journal of Operational Research*, ISSN 0377-2217, 170(3): 909–922.

[23] Miettinen, K., Mäkelä, M. M., Kaario, K. (2006); Experiments with classification-based scalarizing functions in interactive multiobjective optimization. *European Journal of Operational Research*, ISSN 0377-2217, 175(2): 931–947.

[24] Miettinen, K., Mustajoki, J., Stewart, T. J. (2014b); Interactive multiobjective optimization with NIMBUS for decision making under uncertainty. *OR Spectrum*, ISSN 0171-6468, 36(1): 39–56.

[25] Miettinen, K., Ruiz, F., Wierzbicki, A. (2008); Introduction to multiobjective optimization: Interactive approaches, in Branke, J., Deb, K., Miettinen, K., Słowiński, R., editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches (Lecture Notes in Computer Science, 5252)*, Springer, ISBN 978-3-540-88908-3, 27–57.

[26] Narula, S. C., Weistroffer, H. (1989); A flexible method for nonlinear multicriteria decision-making problems. *IEEE Transactions on Systems, Man and Cybernetics*, ISSN 1083-4419, 19(4): 883–887.

[27] Naud, A., Duch, W. (2000); Interactive data exploration using MDS mapping, in *Proceedings of the Fifth Conference: Neural Networks and Soft Computing*, ISBN 83-908587-2-X, 255–260.

[28] Ojalehto, V., Miettinen, K., Laukkanen, T. (2014); Implementation aspects of interactive multiobjective optimization for modeling environments: the case of GAMS-NIMBUS. *Computational Optimization and Applications*, ISSN 0926-6003, 58(3): 757–779.

[29] Ojalehto, V., Miettinen, K., Mäkelä, M. (2007); Interactive software for multiobjective optimization: IND-NIMBUS, *WSEAS Transactions on Computers*, ISSN 2224-2872, 6(1): 87–94.

[30] Petkus, T., Filatovas, E., Kurasova, O. (2009); Investigation of human factors while solving multiple criteria optimization problems in computer network. *Technological and Economic Development of Economy*, ISSN 2029-4913, 15(3): 464–479.

[31] Pohlheim, H. (2006); Multidimensional scaling for evolutionary algorithms – visualization of the path through search space and solution space using Sammon mapping. *Artificial Life*, ISSN 1064-5462, 12(2): 203–209.

[32] Saaty, T., Ozdemir, M. (2003); Why the magic number seven plus or minus two, *Mathematical and Computer Modelling*, ISSN 0895-7177, 38(3–4): 33–244.

[33] Steuer, R. E. (1989); The Tchebycheff procedure of interactive multiple objective programming, in Karpak, B., Zionts, S., editors, *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers (NATO ASI Series, 56)*, Springer, ISBN 978-3-642-74919-3, 235–249.

[34] Valdés, J. J., Barton, A. J. (2007); Visualizing high dimensional objective spaces for multi-objective optimization: A virtual reality approach, in *Proceedings of the IEEE Congress on Evolutionary Computation(CEC 2007)*, IEEE, ISBN 978-1-4244-1339-3, 4199–4206.

[35] Walker, D. J., Everson, R. M., Fieldsend, J. E. (2013); Visualizing mutually nondominating solution sets in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, ISSN 1089-778X, 17(2): 165–184.

[36] Yoshimi, M., Kuhara, T., Nishimoto, K., Miki, M., Hiroyasu, T. (2012). Visualization of pareto solutions by spherical self-organizing map and it's acceleration on a GPU. *Journal of Software Engineering & Applications*, ISSN 1945-3116, 5(3): 129–137.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):520-538, August, 2015.

Game Theoretic Distributed Power Control Algorithms for Uplink Wireless Data in Flat Fading Channels

M. Hayajneh, C. Abdallah

Mohammad Hayajneh*
United Arab Emirates University
College of IT , AlAin, P.O.Box 17551
*Corresponding author: Mhayajneh@uaeu.ac.ae

Chaouki Abdallah
University of New Mexico
Department of Electrical & Computer Engineering
MSC01 1100, 1, Albuquerque, NM 87131-0001, USA
chaouki@ece.unm.edu

Abstract: In this paper we present a game-theoretic power control algorithms for wireless data in CDMA cellular systems under two realistic channels: **(a1)** Fast flat fading channel and **(a2)** Slow flat fading channel. The fading coefficients under both **(a1)** and **(a2)** are studied for three appropriate small scale channel models that are used in the CDMA cellular systems: Rayleigh channel, Rician channel and Nakagami channel. This work is inspired by the results presented by [1] under non-fading channels. In other words, we study the impact of the realistic channel models on the findings in [1] through the followings: we evaluate the average utility function, the average number of bits received correctly at the receiver per one Joule expended, for each channel model. Then, using the average utility function we study the existence, uniqueness of Nash equilibrium (NE) if it exists, and the social desirability of NE in the Pareto sense. Results show that in a non-cooperative game (NPG) the best policy for all users in the cell is to target a fixed signal-to-interference and noise ratio (SINR) similar to what was shown in [1] for non-fading channel. The difference however is that the target SINR in fading channels is much higher than that in a non-fading channel. Also, for spreading gain less than or equal to 100, both NPG and non-cooperative power control game with pricing (NPGP) perform poorly, where all the terminals except the nearest one were not able to attain their corresponding minimum SINR even if sending at the maximum powers in their strategy spaces.
Keywords: Code-division-multiple-access (CDMA), utility function, power control, game theory, non-cooperative game (NPG), wireless data.

1 Introduction

The mathematical theory of games was introduced by Von Neumann and Morgenstern in 1944 [18], and by the late 1970's became an important tool whenever a player's decision depends on what the other players did or will do. A core idea of game theory is how strategic interactions between rational agents (players) generate outcomes according to the players' utilities [10], [19]. Game theory thus forms a suitable framework to obtain more insight into the interactions of self-interested rational agents with potentially conflicting interests. A player in a non-cooperative game responds to the actions of other players by choosing a strategy (from his strategy space) in an attempt to maximize a utility function that quantifies its level of satisfaction.

In a cellular system each user desires to have a high SINR at the base station (BS) coupled with the lowest possible transmit power. It is important in such systems to have high SINR, as this will reflect a low error rate, a more reliable system, and high channel capacity, so that more users can be served per cell [11]. It is also important to decrease the transmit power to lengthen

battery life and to alleviate the near-far problem [17]. In power control algorithms exploiting game theory however, the tendency of each user to maximize his/her utility function in response to other users' actions, leads to a sequence of power vectors that converges to a point where no user has incentive to individually increase his power. This operating point is called a Nash equilibrium. Due to the lack of cooperation between the users this point may not be efficient, in the sense that it may not be the most desirable social point [3]. In Pareto sense, the most desirable social point is actually the power vector that Pareto dominates all other power vectors.

It should be noted that extensive work has been done on non-game theoretic power control algorithms for wireless data and multimedia CDMA cellular networks, e.g. [9, 10, 15, 16]. The power control problem for wireless data CDMA systems was first addressed in the game theoretic framework in [1]- [8]. In this paper the work in [1], which only dealt with deterministic (non-fading) channels, is extended to a realistic wireless CDMA channels by considering the following cases of fading models: A Rayleigh fast/slow flat fading channel model, a Rician[1] fast/slow flat fading channel model and a Nakagamifast/slow flat fading channel model. Where we use the same utility function and evaluate its average in the fading channels mentioned above, then we use these averaged utility functions to study the existence, uniqueness and social desirability of NE operating point under each channel model.

The remaining of this paper is organized as follows: In section 2 we present the utility function and the system model studied in this paper. In section 3 we evaluate the performance of the system for the channel models mentioned above. Non-cooperative power control game (NPG) and Non-cooperative power control game with pricing (NPGP) are discussed briefly in sections 4 and 5, respectively. We then point out the constraints on the new modified strategy spaces to guarantee the existence and uniqueness of Nash equilibrium points for NPG and NPGP under the assumed channel models in section 6. Simulation results are outlined in section 7, and our conclusions are given in section 8.

2 Utility Function and System Model

In general utility functions are used to quantify the satisfaction level a player achieves by choosing an action from its strategy profile, given the other players' actions. A utility function thus maps the player's preferences onto the real line. A formal definition of a utility function may be found in [10].

In a CDMA cellular system, a number of users sharing the spectrum and air interface. Henceforth, each user's transmission adds to the interference of all users at the BS. Each user desires to achieve a high quality of reception at the BS, i.e., a high SINR, while using the minimum possible amount of power in order to extend the battery's life. The conflicting goal of each user to have a high SINR at the BS makes the game theoretic framework suitable for studying and solving the problem.

In this paper we consider the same system model and the same utility function of [1]: Uplink single-cell direct sequence code division multiple access (DS-CDMA) system with N users, where each user transmits frames (packets) of M bits with L information bits. The rate of transmission is R bits/sec for all users. Let P_c represent the *average* probability of correct reception of a frame at the BS, and let p represent the average transmit power level. The utility function for a CDMA system is given by:

$$u = \frac{L\,R}{M\,p}\,f(\gamma) \tag{1}$$

[1]For space limitation, we omit the findings related to Rician channel model. The reader can find these findings in the supplementary document [22].

where $f(\gamma)$ is an efficiency function that approximates P_c, u thus represents the number of information bits successfully received at the BS per joule of consumed energy. With the assumption of no error correction, and *random* packet correct reception rate $\tilde{P}_c$, i.e., $P_c = E\{\tilde{P}_c\}$, is then given by $\prod_{l=1}^{M}(1-\tilde{P}_e(l))$, where $\tilde{P}_e(l)$ is the *random* bit error rate (BER) of the lth bit at a given SINR γ_i. P_e is the average BER, that is $P_e = E\{\tilde{P}_e\}$ (c. f. (9) and (23)). We are assuming that all users in a cell are using non-coherent binary frequency shift Keying (BFSK) modulation, and are transmitting at the same rate R. It should be noted that the efficiency function $f(\gamma)$ has the same expression of P_c in terms of $\tilde{P}_e$, except that $\tilde{P}_e$ is replaced by $2\tilde{P}_e$ [1].

3 Evaluation of The Performance

In this section we find closed-form formulas of the average utility functions under the six assumed channel models. We then use these formulas to study the existence and uniqueness of Nash equilibrium point in section 6. The SINR γ_i at the receiver for the ith user is assumed to be large ($\gamma_i \gg 1$) to combat the fading effect, it is given by [13]:

$$\gamma_i = \frac{W}{R}\frac{p_i\, h_i\, \alpha_i^2}{\sum_{k\neq i}^{N} p_k\, h_k\, \alpha_k^2 + \sigma^2} \tag{2}$$

Where α_i is the path fading coefficient between ith user and the BS and is constant for each bit in a fast flat fading channels **(a1)**, while it is constant for each packet in a slow flat fading channels **(a2)**. W is the spread spectrum bandwidth, p_k is the transmitted power of the kth user, h_k is the path gain between the BS and the kth user, and σ^2 is the variance of the AWGN (additive-white-gaussian-noise) representing the background thermal noise in the receiver. For simplicity we express the interference from all other users as x_i, i.e.

$$x_i = \sum_{k\neq i}^{N} p_k\, h_k\, \alpha_k^2 \tag{3}$$

therefore (2) can be written as:

$$\gamma_i = \gamma_i(\alpha_i, x_i) = \frac{W}{R}\frac{p_i\, h_i}{x_i + \sigma^2}\ \ \alpha_i^2 = \gamma_i'\alpha_i^2 \tag{4}$$

For a given α_i and x_i, the BER, $\tilde{P}(e|\alpha_i, x_i)$, of the ith user using BFSK is given by [13]:

$$\tilde{P}(e|\alpha_i, x_i) = \frac{1}{2}\, e^{-\frac{\gamma_i(\alpha_i, x_i)}{2}} \tag{5}$$

The average BER and average utility functions for this modulation scheme is evaluated next under the previously mentioned channel models.

3.1 Rayleigh Flat Fading Channel

In this case α_i is modelled as a Rayleigh random variable with a probability distribution given by:

$$f^{\alpha_i}(\omega) = \frac{\omega}{\sigma_r^2}\, e^{-(1/2\sigma_r^2)\omega^2}, \quad i = 1, 2, \cdots, N \tag{6}$$

Where $\sigma_r^2 = E\{\alpha_i^2\}/2$ is the measure of the spread of the distribution. In all following calculations, and as a consequence of the multiplicative effect of small and large scale models, it is assumed that $\sigma_r^2 = 1/2$. Using (4) and (6) the distribution of γ_i for a given x_i becomes: $f^{\gamma_i|x_i}(\omega) = \frac{1}{\gamma_i'}\, e^{-(\frac{1}{\gamma_i'})\omega}$

Rayleigh Fast Flat Fading Channel

For the lth bit in the frame, we can rewrite the SINR (4) and the interference (3) for the ith user as follows: $\gamma_i(l) = \frac{W}{R}\frac{p_i\, h_i\, \alpha_i^2(l)}{x_i(l)+\sigma^2}$, and $x_i(l) = \sum_{k\neq i}^{N} p_k\, h_k\, \alpha_k^2(l)$. In this paper, we assume that the fading channel is fast enough to have the fading coefficients $\{\alpha_i(l)\}_{l=1}^{M}$ statistically independent, that is iid (identical independent distributed) random variables. Henceforth, the *averaged* correct reception P_c is given as $(1-P_e)^M$, while the efficiency function $f(\gamma_i)$ is given as $(1-2P_e)^M$, where P_e is *averaged* BER for each bit in the frame, that is $P_e = E\{\tilde{P}_e\}$. We will calculate the averaged P_e next. We can find the conditioned error probability $\tilde{P}(e|x_i)$ by taking the average of (5) with respect to $f^{\gamma_i|x_i}(\omega)$:

$$\begin{aligned}\tilde{P}(e|x_i) &= E\left\{\tilde{P}(e|\gamma_i, x_i)\right\} = \int_0^\infty \tilde{P}(e|\omega, x_i)\, f^{\gamma_i|x_i}(\omega) d\omega \\ &= \frac{1}{2\gamma_i'}\int_0^\infty e^{-(\frac{2+\gamma_i'}{2\gamma_i'})\omega} d\omega = \frac{1}{2+\gamma_i'} \end{aligned} \tag{7}$$

Notice that we dropped the bit index l because the average BER does not depend on l. For large SINR, (7) behaves like:

$$\tilde{P}(e|x_i) \approx \frac{1}{\gamma_i'} = \frac{x_i+\sigma^2}{\frac{W}{R}\, p_i\, h_i} \tag{8}$$

Now, we can find the *averaged* BER P_e by taking the expectation of (8):

$$P_e = E\left\{\tilde{P}(e|x_i)\right\} = \frac{E\{x_i\}+\sigma^2}{\frac{W}{R}\, p_i\, h_i} = \frac{1}{\overline{\gamma_i}} \tag{9}$$

where $\overline{\gamma_i}$ is the ratio of the mean of the received power from user i to the mean of the interference at the receiver and given by:

$$\overline{\gamma_i} = \frac{W}{R}\frac{p_i\, h_i}{\sum_{k\neq i}^{N} p_k\, h_k + \sigma^2} \tag{10}$$

Therefore, the average utility function of the ith user is given by:

$$\boxed{u_i = \frac{L}{M}\frac{R}{p_i}\,(1-\frac{2}{\overline{\gamma_i}})^M} \tag{11}$$

Rayleigh Slow Flat Fading Channel

In a slow flat fading channel model, α_i is assumed to be constant for each packet/frame. The averaged efficiency function $f(\gamma_i)$ is therefore given as the expectation of $(1-2\tilde{P}(e|\alpha_i, x_i))^M$ with respect to the random variables α_i and x_i. One can evaluate $u_i(\mathrm{p}|x_i)$ as follows:

$$\begin{aligned} u_i(\mathrm{p}|x_i) &= \int_0^\infty u_i(\mathrm{p}|\omega, x_i)\, f^{\gamma_i|x_i}(\omega)\, d\omega = \int_0^\infty \frac{L}{M}\frac{R}{p_i}\,(1-e^{-\omega/2})^M\,\frac{1}{\gamma_i'}\, e^{-(\frac{1}{\gamma_i'})\omega}\, d\omega \\ &= \frac{L}{M}\frac{R}{p_i\,\gamma_i'}\sum_{k=0}^{M}(-1)^k\binom{M}{k}\times\int_0^\infty e^{-(\frac{k}{2}+\frac{1}{\gamma_i'})\omega}\, d\omega = \frac{L}{M}\frac{R}{p_i}\sum_{k=0}^{M}\binom{M}{k}\,\frac{2\,(-1)^k}{k\,\gamma_i'+2} \end{aligned} \tag{12}$$

For large SINR ($\gamma_i' \gg 1$), (12) can be approximated by:

$$u(\mathrm{p}|x_i) \approx \frac{L}{M}\frac{R}{p_i}\left(1+\frac{1}{\gamma_i'}\sum_{k=1}^{M}\binom{M}{k}\,\frac{2\,(-1)^k}{k}\right) \tag{13}$$

Averaging (13) with resect to x_i we obtain the average utility function for high SINR:

$$\begin{aligned} u_i &= E\left\{u_i(\mathrm{p}|x_i)\right\} \approx \frac{L\ R}{M\ p_i}\ \left(1+\frac{E\{x_i\}+\sigma^2}{\frac{W}{R}\,p_i\ h_i}\ \sum_{k=1}^{M}\binom{M}{k}\ \frac{2\,(-1)^k}{k}\right) \\ &= \frac{L\ R}{M\ p_i}\ \left(1+\frac{1}{\overline{\gamma_i}}\ \sum_{k=1}^{M}\binom{M}{k}\ \frac{2\,(-1)^k}{k}\right) \\ &\boxed{u_i\ \approx\ \frac{L\ R}{M\ p_i}\ \left(1-\frac{\beta}{\overline{\gamma_i}}\right)} \end{aligned} \tag{14}$$

where $\beta = -\sum_{k=1}^{M}\binom{M}{k}\ \frac{2\,(-1)^k}{k} > 0$.

3.2 Nakagami Flat Fading Channel

Here, the fading coefficient α_i is modelled as a Nakagami random variable with a probability distribution given by [13]: $f^{\alpha_i}(\omega) = \frac{2m^m}{\Gamma(m)\Omega^m}\,\omega^{2m-1}\,e^{(-\frac{m}{\Omega})\omega^2};\ \ i=1,2,\cdots,N$ where $\Omega = E\{\alpha_i^2\}$ controls the spread of the distribution. The fading figure $m = \frac{\Omega^2}{E\{(\alpha_i^2-\Omega)^2\}}$ is a measure of the severity of the fading channel, where $m=\infty$ corresponds to a nonfading channel. In the following it is assumed that $\Omega = 1$. Then the distribution of γ_i for fixed x_i is given as: $f^{\gamma_i|x_i}(\omega) = \frac{1}{\Gamma(m)}\ \left(\frac{m}{\gamma_i'}\right)^m\ \omega^{m-1}\ e^{-(\frac{m}{\gamma_i'})\omega}$

Nakagami Fast Flat Fading Channel

We find the conditioned error probability $\tilde{P}(e|x_i)$ as:

$$\begin{aligned} \tilde{P}(e|x_i) &= \int_0^\infty \tilde{P}(e|\omega,x_i)\,f^{\gamma_i|x_i}(\omega)d\omega = \frac{1}{2\Gamma(m)}\ \left(\frac{m}{\gamma_i'}\right)^m \int_0^\infty \omega^{m-1}\ e^{-(\frac{\gamma_i'+2m}{2\gamma_i'})\,\omega}d\omega \\ &= \frac{1}{2}\left(\frac{2m}{2m+\gamma_i'}\right)^m \end{aligned} \tag{15}$$

For fixed m and $\gamma_i' \gg 1$, (15) can be rewritten as:

$$\tilde{P}(e|x_i) \approx \frac{1}{2}\left(\frac{2m}{\gamma_i'}\right)^m \tag{16}$$

To find the *average* P_e, we need to find the mean of $(x_i+\sigma^2)^m$. Here, x_i is a summation of independent random variables each distributed according to a Gamma density function. This makes the evaluation of $(x_i+\sigma^2)^m$ tedious and it may be easier to find an approximate density function of x_i. To do this, let us recall Esseen's inequality which estimates the deviation of the exact distribution of a sum of independent variables from the normal distribution [21].

Theorem 1. *let $Y_1,\cdots,Y_N$ be independent random variables with $EY_j=0$, $E|Y_j|^3<\infty$ $(j=1,\cdots,N)$. Let $\sigma_j^2 = EY_j^2$, $B_N=\sum_{j=1}^N\sigma_j^2$, $L_N = B_N^{-3/2}\sum_{j=1}^N E|Y_j|^3$. Let $\psi_N(z)$ be the c.f. (cumulative distribution) of the random variable $B_N^{-1/2}\sum_{j=1}^N Y_j$. Then*

$$|\psi_N(z) - e^{-z^2/2}| \le 16\,L_N\ |z|^3\ e^{-z^2/3} \tag{17}$$

Define $\tilde{Y}_k = p_k h_k \alpha_k^2$ and $Y_k = \tilde{Y}_k - p_k h_k$. By simple calculations we can find that $\tilde{Y}_k, (k = 1, \cdots, N)$ are Gamma distributed random variables, such that $f^{\tilde{Y}_k}(\omega) = \frac{(m/p_k h_k)^m}{\Gamma(m)}\ \omega^{m-1}\, e^{-(m/p_k h_k)\omega}$ and $E\tilde{Y}_k = p_k h_k$, which means that $Y_k, (k = 1, \cdots, N)$ are zero mean random variables. Note that $\sigma_k^2 = EY_k^2 = (p_k h_k)^2/m, \forall k = 1, \cdots, N$, and therefore, $B_N = \frac{1}{m}\sum_{k=1}^{N}(p_k h_k)^2$. It is fairly simple to find out that the third moment $E|Y_k|^3 = EY_k^3 = \frac{2(p_k h_k)^3}{m^2}$ $(Y_k \geq 0)$, and $L_N = \frac{2\sum_{k=1}^{N}(p_k h_k)^3}{\sqrt{m}\,(\sum_{k=1}^{N}(p_k h_k)^2)^{3/2}}$. For large N, L_N has a very small value, i.e., $L_N << 1$. Examining (17) for small values of z, L_N takes care of the bound and making it very small, while for large values of z, the exponential will decrease the bound and make it approach zero. In conclusion, we can approximate x_i as a Gaussian random variable with mean ζ_{x_i} and variance $\overline{\sigma}_{x_i}^2$ given by:

$$\zeta_{x_i} = E\{x_i\} = E\left\{\sum_{k\neq i}^{N} \alpha_k^2 p_k\, h_k\right\} = \sum_{k\neq i}^{N} p_k\, h_k\, E\{\alpha_k^2\} = \sum_{k\neq i}^{N} p_k\, h_k \tag{18}$$

and

$$\overline{\sigma}_{x_i}^2 = E\{x_i^2\} - \zeta_{x_i}^2 = E\left\{\sum_{l\neq i}^{N}\sum_{k\neq i}^{N} p_l h_l p_k\, h_k\, \alpha_l^2\, \alpha_k^2\right\} - \zeta_{x_i}^2 = \frac{1}{m}\ \sum_{k\neq i}^{N}(p_k\, h_k)^2 \tag{19}$$

where (19) was obtained using the fact that α_k and α_l are statistically independent for all $k \neq l$. So, we can write f^{x_i}, the PDF of x_i, as follows: $f^{x_i}(w) = \frac{\delta_i}{\sqrt{2\pi}\overline{\sigma}_{x_i}}\, e^{-\frac{(w-\zeta_{x_i})^2}{2\overline{\sigma}_{x_i}^2}}$, where $w \geq 0$ and $\delta_i = 2/(1 + Erf[\zeta_{x_i}/\sqrt{2}\,\sigma_{x_i}])$ is a scaling factor such that $f^{x_i}(w)$ is a valid PDF. $Erf[.]$ is the error function. By examining equations (18) and (19), one can see that $\zeta_{x_i} \gg \sigma_{x_i}$, therefore $\delta_i \approx 1$. Averaging (16) over $f^{x_i}(\omega)$ we obtain the average error probability P_e for high SINR below:

$$\begin{aligned}
P_e &\approx \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m \int_0^{\infty} \left(x_i + \sigma^2\right)^m \times \frac{1}{\sqrt{2\pi}\overline{\sigma}_{x_i}}\ e^{-\frac{(x_i-\zeta_{x_i})^2}{2\overline{\sigma}_{x_i}^2}}\, dx_i \\
&= \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m \int_{\sigma^2}^{\infty} y^m \times \frac{1}{\sqrt{2\pi}\overline{\sigma}_{x_i}}\ e^{-\frac{(y-(\zeta_{x_i}+\sigma^2))^2}{2\overline{\sigma}_{x_i}^2}}\, dy \\
&\approx \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m \int_0^{\infty} y^m \times \frac{1}{\sqrt{2\pi}\overline{\sigma}_{x_i}}\ e^{-\frac{(y-(\zeta_{x_i}+\sigma^2))^2}{2\overline{\sigma}_{x_i}^2}}\, dy
\end{aligned} \tag{20}$$

where we used the change of variable $y = x_i + \sigma^2$ and the last approximation in (20) used the fact that $\sigma^2 \ll 1$. By examining (20) one can see that it is the mth moment of a random variable y normally distributed with mean $\zeta_y = \zeta_{x_i} + \sigma^2$ and variance $\sigma_y^2 = \overline{\sigma}_{x_i}^2$. Therefore, the average P_e is given by:

$$\begin{aligned}
P_e &= \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m E\left\{y^m\right\} = \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m E\left\{((y-\zeta_y)+\zeta_y)^m\right\} \\
&= \frac{1}{2}\left(\frac{2m}{\frac{W}{R}\, p_i\, h_i}\right)^m \sum_{k=0}^{m}\binom{m}{k}\, \zeta_y^{m-k}\, C_k = \frac{1}{2}\left(\frac{2\,m\,\zeta_y}{\frac{W}{R}\, p_i\, h_i}\right)^m \sum_{k=0}^{m}\binom{m}{k}\, \frac{C_k}{\mu_y^k} \\
&= 2^{m-1}\left(\frac{m}{\overline{\gamma_i}}\right)^m \sum_{k=0}^{m}\binom{m}{k}\, \frac{C_k}{\zeta_y^k}
\end{aligned} \tag{21}$$

where $\overline{\gamma_i}$ is given in (10), and C_k is the kth central moment and it is given by [13]:

$$C_k = \begin{cases} 1.3 \cdots (k-1)\, \overline{\sigma}_{x_i}^k & k \text{ even} \\ 0 & k \text{ odd} \end{cases}$$

By splitting up the summation in (21), we obtain:

$$\sum_{l=0}^{m} \binom{m}{l} \frac{C_l}{\zeta_y^l} = 1 + \binom{m}{2} \frac{\overline{\sigma}_{x_i}^2}{(\sigma^2 + \sum_{k \neq i}^{N} p_k\, h_k)^2} + \cdots + \binom{m}{m'} \frac{1.3 \cdots (m'-1) \overline{\sigma}_{x_i}^{m'-1}}{(\sigma^2 + \sum_{k \neq i}^{N} p_k\, h_k)^{m'}} \tag{22}$$

where $m' = m$ if m is even and $m' = m - 1$ if m is odd. Since $\overline{\sigma}_x^2$ (see (19)) is very small compared to ζ_{x_i} (see (18)), we can approximate the summation by its leading term which is 1. Therefore the average P_e at high SINR behaves like:

$$P_e \approx 2^{m-1} \left(\frac{m}{\overline{\gamma_i}} \right)^m \tag{23}$$

And the average utility function of the ith user is given by:

$$\boxed{u_i = \frac{L\ R}{M\ p_i} \left(1 - 2^m \left(\frac{m}{\overline{\gamma_i}} \right)^m \right)^M} \tag{24}$$

Notice that if we set $m = 1$, we obtain the same performance as in the Rayleigh fast flat fading case.

Nakagami Slow Flat Fading Channel

As done earlier, $u_i(\mathrm{p}|x_i)$ can be determined as follows:

$$\begin{aligned} u_i(\mathrm{p}|x_i) &= \int_0^\infty u_i(\mathrm{p}|\omega, x_i) f^{\gamma_i|x_i}(\omega) d\omega \\ &= \int_0^\infty \frac{L\ R}{M\ p_i} (1 - e^{-\omega/2})^M \frac{1}{\Gamma(m)} \left(\frac{m}{\gamma_i'} \right)^m \omega^{m-1} e^{-(\frac{m}{\gamma_i'})\omega} d\omega \end{aligned} \tag{25}$$

By factorizing $(1 - e^{-\gamma_i/2})^M$ and using the identity $\int_0^\infty y^n e^{-a\,y}\, dy = \frac{\Gamma(n+1)}{a^{n+1}}$ we obtain:

$$u_i(\mathrm{p}|x_i) = \frac{L\,R}{M\,p_i} \sum_{k=0}^{M} (-1)^k \binom{M}{k} \left(\frac{2\,m}{k\,\gamma_i' + 2\,m} \right)^m \tag{26}$$

For fixed m and high SINR, $\gamma_i' \gg 1$ (26) can be approximated as:

$$u_i(\mathrm{p}|x_i) \approx \frac{L\,R}{M\,p_i} \left[1 + (\frac{1}{\gamma_i'})^m \sum_{k=1}^{M} (-1)^k \binom{M}{k} \left(\frac{2\,m}{k} \right)^m \right] \tag{27}$$

Averaging (27) with respect to the distribution of x_i and using the same argument as in (20), (21) and (22) we end up with the final approximate averaged utility function given by:

$$\begin{aligned} u_i &\approx \frac{L\,R}{M\,p_i} \left[1 + (\frac{1}{\overline{\gamma_i}})^m \sum_{k=1}^{M} (-1)^k \binom{M}{k} \left(\frac{2\,m}{k} \right)^m \right] \\ \boxed{u_i} &\approx \boxed{\frac{L\,R}{M\,p_i} \left[1 - \xi\ (\frac{1}{\overline{\gamma_i}})^m \right]} \end{aligned} \tag{28}$$

where $\xi = -\sum_{k=1}^{M} (-1)^k \binom{M}{k} \left(\frac{2\,m}{k}\right)^m > 0$.
In the following two sections, we introduce briefly both NPG and NPGP games.

4 Non-Cooperative Power Control Game (NPG)

Let $\mathcal{N} = \{1, 2, \cdots, N\}$ be the index set of the users currently served in the cell and $\{P_j\}_{j\in\mathcal{N}}$ represent the set of strategy spaces of all users in the cell. Let $G = [\mathcal{N}, \{P_j\}, \{u_j(.)\}]$ denote a noncooperative game, where each user, based on local information, chooses a power level from a convex set $P_j = [p_{j-min}, p_{j-max}]$ and where p_{j-min} and p_{j-max} are the minimum and the maximum power levels in the jth user strategy space, respectively. Assuming that the power vector $\mathrm{p} = [p_1, p_2, \cdots, p_N]$ is the result of NPG, the utility of user j is given by [1]:

$$u_j(\mathrm{p}) = u_j(p_j, \mathrm{p}_{-j}) \tag{29}$$

where p_j is the power of user j, and p_{-j} is the vector of powers transmitted by all other users. The right side of (29) emphasizes the fact that user j can only control his own power. We rewrite (1) for user j as:

$$u_j(p_j, \mathrm{p}_{-j}) = \frac{L\,R}{M\,p_j}\,f(\gamma_j) \tag{30}$$

The formal expression for the NPG is given in [1] as:

$$G: \quad \max_{p_j\in \mathrm{P}_j} u_j(p_j, \mathrm{p}_{-j}), \ \text{ for all } \ j \in \mathcal{N} \tag{31}$$

This game will produce a sequence of power vectors until it converges to a point where all users are satisfied with their utility level. This operating point is called a Nash equilibrium operating point of NPG. In the next subsection, we define the Nash equilibrium point and describe its physical interpretation.

4.1 Nash Equilibrium in NPG

Definition 2. *[1] A power vector* $\mathrm{p} = [p_1, p_2, \cdots, p_N]$ *is a Nash equilibrium of the* NPG defined above *if for every* $j \in \mathcal{N}, u_j(p_j, \mathrm{p}_{-j}) \geq u_j(p'_j, \mathrm{p}_{-j})$ *for all* $p'_j \in \mathrm{P}_j$.

One interpretation of Nash equilibrium is that no user can increase its utility by changing its power level unilaterally. If we multiply the power vector p by a constant $0 < \lambda < 1$ we may get higher utilities for all users, as was the case in nonfading channels. This means that the Nash equilibrium is not efficient, that is, the resulting p is not the most desired social operating point. This results from the lack of cooperation between the users currently in the system. To impose a level of cooperation between users in order to reach a Pareto dominant Nash point, a pricing technique was introduced in [1]. We discus this modified NPG game next.

5 Non-Cooperative Power Control Game with Pricing (NPGP)

In NPGP each user maximizes the difference between his/her utility function and a pricing function. This aims to allow more efficient use of the system resources within the cell, as each user is made aware of the cost of aggressive resources use, and of the harm done to other users in the cell. We use here a linear pricing function, i.e., a pricing factor multiplied by the transmit power. The base station broadcasts the pricing factor to help the users currently in the cell reach a Nash equilibrium that improves the aggregate utilities of all users at power levels lower than those of the pure NPG. In other words, the resulting power vector of NPGP is Pareto dominant compared to the resulting power vector of NPG, but is still not Pareto optimal in the sense that we may multiply the resulting power vector of NPGP by a constant $0 < \lambda < 1$ to get higher

utilities for all users. Let $G_c = [\mathcal{N}, \{\mathrm{P}_j\}, \{u_j^c(.)\}]$ represent an N-player noncooperative power control game with pricing (NPGP), where the utilities are [1]:

$$u_j^c(\mathrm{p}) = u_j(\mathrm{p}) - c\,p_j \text{ for all } j \in \mathcal{N} \tag{32}$$

where c is a positive number chosen to get the best possible improvement in the performance. Therefore, NPGP with a linear pricing function can be expressed as:

$$G_c: \quad \max_{p_j \in \mathrm{P}_j} \{u_j(\mathrm{p}) - c\,p_j\} \text{ for all } j \in \mathcal{N} \tag{33}$$

6 Existence and Uniqueness of Nash Equilibrium Point

In this section we show that NPG and NPGP introduced by [1] admit a unique Nash equilibrium points under the assumed channel models. However, to guarantee the existence and uniqueness of NE point in both games, the terminals' strategy spaces defined in [1] should be constrained more. That is, some transmit power values which were allowed in a non-fading channel, may not be allowed under a fading channel. In the following, we refer to the unconstrained maximizing transmit power level of user i by p_i^{max}. P_i refers to the convex strategy space of user i.

Lemma 3. *In NPG under Rayleigh fast flat fading channel with the average utility function u_i given in (11), the existence of a Nash equilibrium point is guaranteed if and only if the strategy space is modified to $P_i = \{p_i : \overline{\gamma_i} \in \big(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\big)\}$, where $\overline{\gamma}_{i-min} = 2(M+1) - \sqrt{2M(M+1)}$ and $\overline{\gamma}_{i-max} = 2(M+1) + \sqrt{2M(M+1)}$. The best response vector of all users $r^1(p) = (r_1^1(p), r_2^1(p), \cdots, r_N^1(p))$, where $r_i^1(p) = min(p_i^{max}, p_{i-max})$, and*

$$p_i^{max} \;=\; 2(M+1)\,I_i, \;\; I_i = \frac{R\,(\sum_{k\neq i}^{N} h_k p_k + \sigma^2)}{W\,h_i} \tag{34}$$

is a standard interference function, therefore by [12], the Nash equilibrium point is unique. I_i is the effective interference for user i.

Proof: In all following proofs we make use of the classical results of game theory, where the existence of a Nash equilibrium point is guaranteed if the utility function is quasiconcave and optimized on a convex strategy space. Thus, to prove the existence of Nash equilibrium point, it is enough to prove that the utility function u_i is concave (a concave function on some set is also a quasiconcave function on the same set) in p_i given p_{-i} on the convex set $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \big(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\big)\}$. Let us find the first and second order derivatives with respect to p_i as follows:

$$\frac{\partial u_i}{\partial p_i} = \frac{L\,R}{M\,p_i^2}\,\Big(\frac{2(M+1)}{\overline{\gamma_i}} - 1\Big)\,\Big(1 - \frac{2}{\overline{\gamma_i}}\Big)^{M-1}, \tag{35}$$

then $\frac{\partial^2 u_i}{\partial p_i^2} = \frac{2L\,R\,(1-\frac{2}{\overline{\gamma_i}})^M\,[2(M+1)(2+M) - 4(1+M)\,\overline{\gamma_i} + \overline{\gamma_i}^2]}{M\,p_i^3(-2+\overline{\gamma_i})^2}$. Therefore, $\frac{\partial^2 u_i}{\partial p_i^2} < 0, \forall\ \overline{\gamma_i} \in \big(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\big)$, where $\overline{\gamma}_{i-min} = 2(M+1) - \sqrt{2M(M+1)}$ and $\overline{\gamma}_{i-max} = 2(M+1) + \sqrt{2M(M+1)}$. This implies that the strategy space should be modified to the convex set $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \big(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\big)\}$ to guarantee the concavity of the utility function, and then to guarantee the existence of Nash equilibrium point.

To prove the uniqueness of Nash equilibrium point we need to prove that $r^1(\mathrm{p})$ is a standard function. By setting (35) to zero we find the maximizing transmit power level that lies in the convex strategy space P_i is given as in (34). Before proving that $r^1(\mathrm{p})$ is a standard interference function we introduce the definition of an arbitrary standard interference function $\phi(\mathrm{p})$ as follows:

Definition 4. *[12] An interference function $\phi(p)$ is called a standard interference function if it satisfies the following: 1) Positivity: $\phi(p) > 0$, i.e., each element is positive, 2) Monotonicity: if $p > \hat{p}$ then $\phi(p) \geq \phi(\hat{p})$ (component wise), and 3) Scalability: $\forall \delta > 1,\ \delta\phi(p) > \phi(\delta\, p)$ (component wise).*

To prove that $r^1(\mathrm{p})$ is a standard interference function we proceed as follows: The proof of positivity is trivial, since $\mathrm{P}_i \subset \mathbb{R}^+$ and $r_i^1(\mathrm{p}_{-i}) \in P_i, \forall\, i \in \mathcal{N}$, where $r_i^1(\mathrm{p}_{-i}) = r_i^1(\mathrm{p})$. Also, it is obvious that $p_i^{max}(\mathrm{p}) > p_i^{max}(\hat{\mathrm{p}})$ for all i if $\mathrm{p} > \hat{\mathrm{p}}$ by looking at (34), henceforth the monotonicity of $r^1(\mathrm{p})$ is satisfied. To prove the scalability, it is enough to prove that $p_i^{max}(\mathrm{p}_{-i})$ is a scalable function and then the scalability of $r^1(\mathrm{p})$ comes through. Let us rewrite equation (34) as follows: $p_i^{max}(\mathrm{p}_{-i}) = \frac{2R\,(M+1)\,(\sum_{k\neq i}^{N} h_k p_k + \sigma^2)}{W\,h_i}$ then

$$p_i^{max}(\delta \mathrm{p}_{-i}) = \frac{2R\,(M+1)\,(\,\delta \sum_{k\neq i}^{N} h_k p_k + \sigma^2)}{W\,h_i}, \tag{36}$$

while

$$\delta p_i^{max}(\mathrm{p}_{-i}) = \frac{2\,\delta\,R\,(M+1)\,(\sum_{k\neq i}^{N} h_k p_k + \sigma^2)}{W\,h_i} \tag{37}$$

It is clear that $\delta p_i^{max}(\mathrm{p}_{-i}) > p_i^{max}(\delta\, \mathrm{p}_{-i})$, therefore $r^1(\mathrm{p})$ is a standard interference function, and the Nash equilibrium point is unique. □

In the following lemmas we omit the proof of existence and/or uniqueness as they are similar to those of lemma 3.

Lemma 5. *In NPG under Rayleigh slow flat fading channel with the average utility function (14), the existence of a Nash equilibrium point is guaranteed if and only if the strategy space is modified to the convex set $P_i = \{p_i : \overline{\gamma_i} \in (1, 3\,\beta)\}$. The best response vector of all users $r^2(p) = (r_1^2(p), r_2^2(p), \cdots, r_N^2(p))$, where $r_i^2(p) = min(p_i^{max}, p_{i-max})$, and $p_i^{max} = 2\,\beta\, I_i$, is a standard interference function, therefore by [12] Nash equilibrium point is unique.*

Proof: Similarly, we need to find the first and second order derivatives of u_i with respect to p_i: $\frac{\partial u_i}{\partial p_i} = \frac{L\,R}{M\,p_i^2}\left(\frac{2\,\beta}{\overline{\gamma_i}} - 1\right)$, then $\frac{\partial^2 u_i}{\partial p_i^2} = \frac{2\,L\,R}{M\,p_i^3}\left(1 - \frac{3\,\beta}{\overline{\gamma_i}}\right)$, which means that $\frac{\partial^2 u_i}{\partial p_i^2} < 0, \forall\, \overline{\gamma_i} \in (1, 3\,\beta)$ so the convex strategy space should have the following form: $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in (1, 3\,\beta)\}$ to guarantee the concavity of u_i and then to guarantee the existence of Nash point. □

Lemma 6. *In NPG under Nakagami fast flat fading channel with the average utility function u_i given in (24) with $m = 2$, the existence of a Nash equilibrium point is guaranteed if the strategy space is modified to the following convex set $P_i = \{p_i : \overline{\gamma_i} \in (\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max})\}$, where $\overline{\gamma}_{i-min} = \sqrt{8}\sqrt{2 + 5M - \sqrt{M\,(8 + 17M)}}$ and $\overline{\gamma}_{i-max} = \sqrt{8}\sqrt{2 + 5M + \sqrt{M\,(8 + 17M)}}$. The best response vector of all users $r^5(p) = (r_1^5(p), r_2^5(p), \cdots, r_N^5(p))$, where $r_i^5(p) = min(p_i^{max}, p_{i-max})$, and $p_i^{max} = 4\sqrt{1 + 2M}\, I_i$, is a standard interference function, therefore by [12] Nash equilibrium point is unique.*

Proof: We find the first and second order derivatives of u_i in (24) after setting $m = 2$ with respect to p_i as follows:

$$\frac{\partial u_i}{\partial p_i} = \frac{L\,R}{M\,p_i^2}\,(\frac{16(2M+1)}{\overline{\gamma}_i^2} - 1)\,(1 - \frac{16}{\overline{\gamma}_i^2})^{M-1}, \tag{38}$$

then

$$\frac{\partial^2 u_i}{\partial p_i^2} = \frac{1}{M\,p_i^3(-16 + \overline{\gamma}_i^2)^2}\left(2\,L\,R\,(1 - \frac{16}{\overline{\gamma}_i^2})^M \times \left[256(1+M)(2M+1) - 16(2+5M)\,\overline{\gamma}_i^2 + \overline{\gamma}_i^4\right]\right) \tag{39}$$

and this implies that $\frac{\partial^2 u_i}{\partial p_i^2} < 0, \forall\ \overline{\gamma_i} \in \left(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\right)$,
where $\overline{\gamma}_{i-min} = \sqrt{8(2+5M) - 8\sqrt{M\,(8+17M)}}$ and $\overline{\gamma}_{i-max} = \sqrt{8(2+5M) + 8\sqrt{M\,(8+17M)}}$. Henceforth, the strategy space should have the following convex set: $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \left(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\right)\}$ to guarantee that u_i is strict concave on P_i, then a Nash equilibrium exists. □

Lemma 7. *In NPG under Nakagami slow flat fading channel with the average utility function u_i given in (28), a Nash equilibrium point is guaranteed if and only if the strategy space is the following convex set $P_i = \{p_i : \overline{\gamma_i} \in \left(1, \sqrt{6\,\xi}\right)\}$. The best response vector of all users $r^6(p) = (r_1^6(p),\, r_2^6(p),\, \cdots,\, r_N^6(p))$, where $r_i^6(p) = min(p_i^{max}, p_{i-max})$, and $p_i^{max} = \sqrt{3\xi}\ I_i$, is a standard interference function, therefore by [12] Nash equilibrium point is unique.*

Proof: The first derivative and second order derivatives of u_i after setting $m = 2$ with respect to p_i are given by: $\frac{\partial u_i}{\partial p_i} = \frac{L\,R}{M\,p_i^2}\left(\frac{3\,\xi}{\overline{\gamma}_i^2} - 1\right)$, and $\frac{\partial^2 u_i}{\partial p_i^2} = \frac{2\,L\,R}{M\,p_i^3}\left(1 - \frac{6\,\xi}{\overline{\gamma}_i^2}\right)$, therefore $\frac{\partial^2 u_i}{\partial p_i^2} < 0, \forall\ \overline{\gamma_i} \in \left(1, \sqrt{6\,\xi}\right)$. As a result, the convex strategy space should be $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \left(1, \sqrt{6\,\xi}\right)\}$ to guarantee the strict concavity of u_i and then the existence of a Nash equilibrium point is guaranteed. □

Now, we turn to the existence and uniqueness of Nash equilibrium point of NPGP under the assumed channel models discussed above.

Lemma 8. *In NPGP under Rayleigh fast flat fading channel model with utility function $u_i^c = u_i - c\ p_i$, where u_i is given in (11), a Nash equilibrium point existence is guaranteed if and only if the strategy space is the following convex set: $P_i = \{p_i : \overline{\gamma_i} \in \left(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\right)\}$, where $\overline{\gamma}_{i-min} = 2\,(M+1) - \sqrt{2\,M(M+1)}$ and $\overline{\gamma}_{i-max} = 2\,(M+1)$. The best response vector of all users $r^7(p) = (r_1^7(p), r_2^7(p), \cdots, r_N^7(p))$, where $r_i^7(p) = min(p_i^{max}, p_{i-max})$, and*

$$p_i^{max} \approx \frac{-6\,2^{1/3}\,a + 2^{2/3}(27\,b_i + \sqrt{108\,a^3 + 729\,b_i^2}\ )^{2/3}}{6\,(27\,b_i + \sqrt{108\,a^3 + 729\,b_i^2}\ )^{1/3}};\ a = \frac{L\,R}{M\,c},\ \ b_i = \frac{2\,(M+1)\,L\,R\,I_i}{M\,c} \quad (40)$$

is a standard interference function, therefore by [12] Nash equilibrium point is unique.

Proof: Let us find the maximizing power p_i^{max} in terms of the SINR $\overline{\gamma_i}$ as follows:

$$\frac{\partial u_i^c}{\partial p_i} = \frac{L\,R}{M\,p_i^2}\,(\frac{2\,(M+1)}{\overline{\gamma_i}} - 1)\,(1 - \frac{2}{\overline{\gamma_i}})^{M-1} - c = 0, \quad (41)$$

then

$$p_i^{max} = \sqrt{\frac{L\,R}{M\,c}\,(\frac{2\,(M+1)}{\overline{\gamma_i}} - 1)\,(1 - \frac{2}{\overline{\gamma_i}})^{M-1}} \quad (42)$$

For p_i^{max} to be feasible, i.e., real and positive we need to satisfy the following condition on the strategy space: $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in (1, M+1)\}$. But, to guarantee the concavity of the utility function u_i^c, we have to have $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \left(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\right)\}$, where $\overline{\gamma}_{i-min} = 2\,(M+1) - \sqrt{2\,M(M+1)}$ and $\overline{\gamma}_{i-max} = 2\,(M+1) + \sqrt{2\,M(M+1)}$. Therefore, to fulfill the two conditions the convex strategy space should be the intersection of the two sets, that is $\mathrm{P}_i = \{p_i : \overline{\gamma_i} \in \left(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\right)\}$, where $\overline{\gamma}_{i-min} = 2\,(M+1) - \sqrt{2\,M(M+1)}$ and $\overline{\gamma}_{i-max} = 2\,(M+1)$. Since $\overline{\gamma_i} >> 1$ on the convex strategy space P_i given above, one can approximate p_i^{max}, the solution of (41), as the feasible solution of the following equation:

$$p_i^3 + \frac{L\,R}{M\,c}\,p_i - \frac{2\,(M+1)\,L\,R\,I_i}{M\,c} = 0 \quad (43)$$

This equation has only one real and positive solution which is given in (40). It is fairly easy to prove that $r^7(\mathrm{p})$ with p_i^{max} as given in (40) is a standard interference function. Therefore, Nash equilibrium point is unique. □

In the following lemmas we may omit the proof of existence and/or uniqueness if it can be argued the same way as in lemma 8.

Lemma 9. *In NPGP under Rayleigh slow flat fading channel model with utility function $u_i^c = u_i - c\ p_i$, where u_i is given in (14), a Nash equilibrium point existence is guaranteed if and only if the strategy space is the following convex set: $P_i = \{p_i : \overline{\gamma_i} \in (1, 2\ \beta)\}$. The best response vector of all users $r^8(p) = (r_1^8(p), r_2^8(p), \cdots, r_N^8(p))$, where $r_i^8(p) = min(p_i^{max}, p_{i-max})$, and*

$$p_i^{max} = \frac{-6\,2^{1/3}\,a + 2^{2/3}(27\,b_i + \sqrt{108\,a^3 + 729\,b_i^2}\ )^{2/3}}{6\,(27\,b_i + \sqrt{108\,a^3 + 729\,b_i^2}\ )^{1/3}};\ a = \frac{L\,R}{M\,c},\ b_i = \frac{2\,L\,R\,\beta\,I_i}{M\,c} \tag{44}$$

is a standard interference function, therefore by [12] Nash equilibrium point is unique. We have equality in (44) because there was no approximation in getting p_i^{max}.

Lemma 10. *In NPGP under Nakagami fast flat fading channel model with utility function $u_i^c = u_i - c\ p_i$, where u_i is given in (24), a Nash equilibrium point existence is guaranteed if and only if the strategy space is the convex set: $P_i = \{p_i : \overline{\gamma_i} \in \big(\overline{\gamma}_{i-min}, \overline{\gamma}_{i-max}\big)\}$, where $\overline{\gamma}_{i-min} = \sqrt{8}\sqrt{2 + 5M - \sqrt{M\,(8 + 17M)}}$ and $\overline{\gamma}_{i-max} = 4\sqrt{1 + 2M}$. The best response vector of all users $r^{11}(p) = (r_1^{11}(p),\ r_2^{11}(p),\ \cdots,\ r_N^{11}(p))$, where $r_i^{11}(p) = min(p_i^{max}, p_{i-max})$, and*

$$p_i^{max} \approx \sqrt{\frac{L\,R}{2\,M\,c}}\sqrt{-1 + \sqrt{1 + \frac{64\,(1 + 2M)\,I_i^2\,M\,c}{L\,R}}} \tag{45}$$

is a standard interference function, therefore by [12] Nash equilibrium point is unique.

Proof: The maximizer transmit power p_i^{max} is the feasible solution of $\frac{\partial\,u_i}{\partial\,p_i} - c = 0$, where $\frac{\partial\,u_i}{\partial\,p_i}$ is given in (38), and results in a polynomial of degree $2M + 4$. It is a tedious and may be impossible to find a closed-form for the feasible solution of this polynomial. Recall that $\overline{\gamma_i} > 4$ to guarantee $u_i(\mathrm{p}) > 0$, so the maximizer transmit power level p_i^{max} can be approximated by the feasible solution of the following equation.

$$p_i^4 + \frac{L\,R}{M\,c}\,p_i^2 - \frac{16\,(1 + 2M)\,L\,R\,I_i^2}{M\,c} = 0 \tag{46}$$

The only feasible solution of the equation above is given by (45). □

Lemma 11. *In NPGP under Nakagami slow flat fading channel model with utility function $u_i^c = u_i - c\ p_i$, where u_i is given in (28), a Nash equilibrium point existence is guaranteed if and only if the strategy space is the following convex set: $P_i = \{p_i : \overline{\gamma_i} \in \big(1, \sqrt{3\,\xi}\big)\}$. The best response vector of all users $r^{12}(p) = (r_1^{12}(p),\ r_2^{12}(p),\ \cdots,\ r_N^{12}(p))$, where $r_i^{12}(p) = min(p_i^{max}, p_{i-max})$, and*

$$p_i^{max} = \sqrt{\frac{L\,R}{2\,M\,c}}\sqrt{-1 + \sqrt{1 + \frac{12\,\xi\,I_i^2\,M\,c}{L\,R}}} \tag{47}$$

is a standard interference function, therefore by [12] Nash equilibrium point is unique.

Proof: The maximizer transmit power level p_i^{max} is the feasible solution of the following equation.

$$p_i^4 + \frac{L\,R}{M\,c}\,p_i^2 - \frac{3\xi\,L\,R\,I_i^2}{M\,c} = 0 \tag{48}$$

The only feasible solution of the equation above is as given by (47). It is simple to check that $r^{12}(\mathrm{p})$ with the maximizer power in (47) satisfies all the conditions of a standard interference function. Henceforth, the Nash equilibrium point is unique. □

Observing lemmas 3-7, we see that the maximizing SINR γ_i^{max} for all users are the same: $\gamma_i^{max} = 2(M+1)$, $\forall\, i \in \mathcal{N}$ under fast Rayleigh and fast Rician fading channels. On the other hand $\gamma_i^{max} = 2\,\beta$, $\forall\, i \in \mathcal{N}$ under slow Rayleigh and slow Rician fading channels. $\gamma_i^{max} = 4\sqrt{1+2M}\,\forall\, i \in \mathcal{N}$ under fast Nakagami fading channels, and $\gamma_i^{max} = \sqrt{3\,\xi}\,\forall\, i \in \mathcal{N}$ under slow Nakagami fading channels. For nonfading channels it was shown in [1] that $\gamma_i^{max} = 12.4$, $\forall\, i \in \mathcal{N}$. This implies, as expected, that in order to overcome the fading effect, users in fading channels have to target higher SINR values.

Next, we introduce an algorithm that converges to Nash equilibrium point of NPG and NPGP. We need to keep in mind that the strategy space denoted by P_i in the algorithm differs according to the channel model. The algorithm is the same as in [1] except that the strategy spaces are modified to the forms given in lemmas 3-7 to guarantee the existence of Nash equilibrium point under the studied channel models.

Assume user j updates its power level at time instances that belong to a set T_j, where $T_j = \{t_{j_1}, t_{j_2}, \cdots\}$, with $t_{j_k} < t_{j_{k+1}}$ and $t_{j_0} = 0$ for all $j \in \mathcal{N}$. Let $T = \{t_1, t_2, \cdots\}$ where $T = T_1 \bigcup T_2 \bigcup \cdots \bigcup T_N$ with $t_k < t_{k+1}$ and define $\underline{\mathrm{p}}$ to be the smallest power vector in the modified strategy space $\mathrm{P} = \mathrm{P}_1 \bigcup \mathrm{P}_2 \bigcup \cdots \bigcup \mathrm{P}_N$.

Algorithm 12. *Consider non-cooperative game G as given in (31) and generate a sequence of power vectors as follows:*
Initialize
$k = 0$;
$N =$ *total numberof active users;*
$p[0] = \underline{p}$;
for $j = 1$ *to* N ***do***
set $T_j =$ *set of times user* j *updates its power;*
end for
set $T = \bigcup_{j \in N} T_j = \{t_1, t_2, t_3, ...\}$;
$k \leftarrow k + 1$;
STEP1 *:* ***for*** $j = 1$ *to* N ***do***
if $t_k \in T_j$;
set $p_j^{max}(t_k) = arg \underset{p_j \in P_j}{max}\; u_j(p_j, p_{-j}(t_{k-1}))$;
else $p_j^{max}(t_k) = p_j^{max}(t_{k-1})$;
end if;
end for
if $p(t_k) = p(t_{k-1})$
stop and declare $p(t_k)$ *as the NE Point;*
else
$k \leftarrow k + 1$
Goto STEP1

The next algorithm finds the best pricing factor c_{Best} for NPGP, keeping in mind that the strategy space should be according to lemmas 8-11.

Algorithm 13. *Initialize*
$\Delta c =$ *real positive number;*
$c = 0$*;*
use algorithm 12 to obtain $u_j^{c=0}$ *for all users* $j = 1$ *to* $j = N$ *;*
STEP2:***for*** $j = 1$ *to* N ***do***
set $c_j \leftarrow c + \Delta c$*; {Broadcast the pricing factor to all users} ;*
end for
use algorithm 12 to obtain $u_j^{c+\Delta c}$ *for all users* $j = 1$ *to* $j = N$ *;*
if $u_j^{c+\Delta c}$ *is Pareto dominates* u_j^c *for all users;*
goto STEP2
else
stop *and declare* $c_{Best} = c - \Delta c$

7 Simulation Results

We show the effects of time-varying, fast and slow fading wireless channels on the equilibrium utilities and powers, that are the outcomes of the extended NPG and NPGP algorithms (algorithms12 and 13) which were originally studied for non-fading wireless channels in [1].

The system studied is a single-cell DS-CDMA cellular mobile system with 9 stationary users, all are using the same data rate R and the same modulation scheme, non-coherent BFSK. The system parameters used in this study are given in Table 1. The distances between the 9 users and the BS are d = [310, 460, 570, 660, 740, 810, 880, 940, 1000] in meters. The path attenuation between user j and the BS using the simple path loss model [17] is $h_j = 0.097/d_j^4$, where 0.097 approximates the shadowing effect. Results of simulations show that under Rayleigh, Rician, and Nakagami fast flat fading channels with spreading gain $W/R = 10^2$, users do not reach a Nash equilibrium point where all users except the nearest user to the BS are using the highest power level in the strategy space without achieving the maximizing SINRs ($\gamma_i^{max} = 2(M+1)$, $\forall i \in \mathcal{N}$).

In Fig.1 we demonstrate the equilibrium utilities and the equilibrium powers of NPG under a fast fading channels **(a1)** with the three small scale fading models with spreading gain $W/R = 10^3$. All users were able to achieve their maximizing SINR under two small scale fading models, namely, Nakagami and Rician channels. Under the Rayleigh channel model however some users failed to achieve the maximizing SINR. One can see in Fig.1 that the farthest 4 users in the Rayleigh channel were forced to send at their maximum allowable power to achieve their minimum SINRs. The equilibrium utilities and equilibrium powers of the NPGP under **(a1)** are shown in the left and right graphs of Fig.2, respectively. Results show that a Pareto improvement over NPG for Rayleigh, Rician, and Nakagami channels was obtained such that all users succeeded to attain SINRs more than their corresponding minimum SINRs ($\gamma_i > \gamma_{i-min}$, $\forall i \in \mathcal{N}$).

Then we present the effect of a slow flat fading channels **(a2)** on the equilibrium utilities and powers which are the outcomes of NPG algorithm 12 as shown in Fig. 3. This figure shows that, unlike fast fading channels, all users succeeded to achieve the maximizing SINR ($\gamma_i^{max} = 2\beta = 19.8619$ under Rician and Rayleigh channels and $\gamma_i^{max} = \sqrt{3\xi} = 25.1182$ under Nakagami channels). Left graph of Fig. 3 shows that under both Rayleigh and Nakagami channel models users were equally satisfied, i.e., the equilibrium utilities are the same for both models. However, under Rayleigh channels, the equilibrium powers are less than those under Nakagami channels. This could be due to the fact that users under Nakagami channels target a higher maximizing SINR as we just mentioned above.

As for the effect of slow fading channels on the outcomes of NPGP, equilibrium utilities and equilibrium powers, our simulations showed that Pareto improvement (dominance) over NPG

was not possible under the three small scale models. At $c = c_{Best}$, simulations showed that the best policy is that all users to target a fixed SINR, that is $\gamma_i^{max} = 19.8619$ under Rician and Rayleigh channels and $\gamma_i^{max} = 25.1182$ under Nakagami channels, which is exactly the same situation as in NPG. To demonstrate this result for the three small scale models more clearly, we present Fig. 4 for Rician and Rayleigh channel models and Fig. 5 for Nakagami channel model. Fig. 4 shows that with $c = c_{Best}$ the maximizing transmit power p_i^{max} given in (44) behaves with respect to the effective interference I_i (feasible values of I_i) the same as $p_i^{max} = 2\,\beta\,I_i$ given in Lemmas 2 and 4. While Fig. 5 shows that p_i^{max} given in (47) behaves with feasible values of I_i the same as $p_i^{max} = \sqrt{3\,\xi}\,I_i$ given in Lemma 6. Surprisingly, both figures suggest that NPGP with linear pricing does not admit a Pareto dominance over the NE operating point of NPG in a slow flat fading channels under the three small scale fading models.

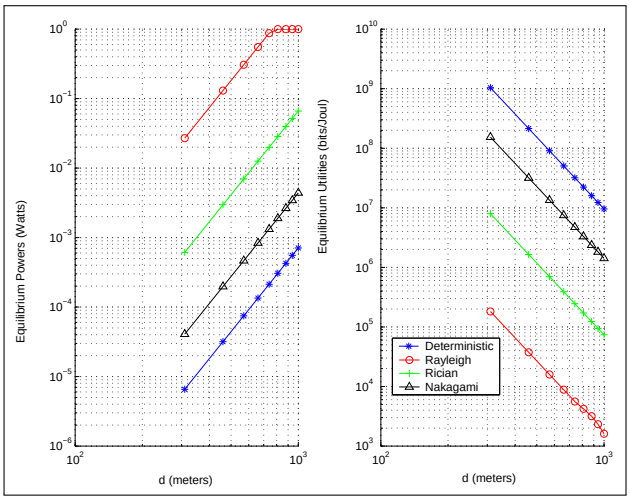


Figure 1: Equilibrium powers and equilibrium utilities of NPG for Rician fast flat fading channel gain (+), Rayleigh fast flat fading channel gain (o), Nakagami fast flat fading ($\triangle$) and deterministic channel gain (*) versus the distance of a user from the BS in meters with $W/R = 10^3$.

Table 1: The values of parameters used in the simulations.

L, number of information bits	64
M length of the codeword	80
W, spread spectrum bandwidth	10^6, 10^7 Hz
R, data rate	10^4 bits/sec
σ^2, AWGN power at the BS	5×10^{-15}
N, number of users in the cell	9
s^2, specular component	1
W/R, spreading gain	10^2, 10^3
m, fading figure	2
p_{i-max}, maximum power in ith user's strategy space	1 Watts

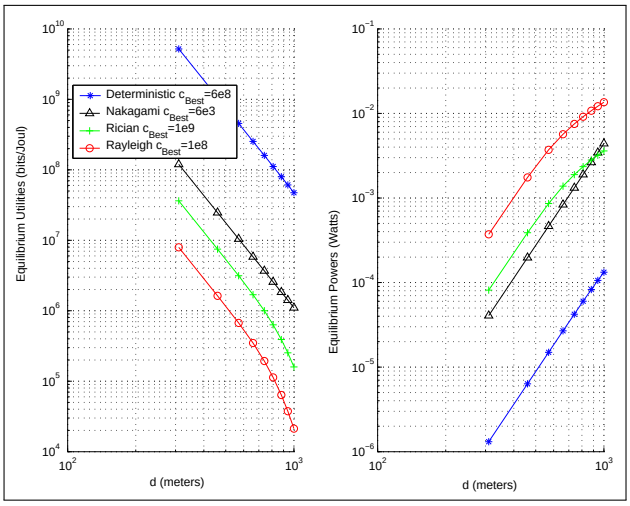


Figure 2: Equilibrium utilities and equilibrium powers of NPGP for Rician fast flat fading channel gain (+), Rayleigh fast flat fading channel gain (o), Nakagami fast flat fading ($\triangle$) and deterministic channel gain (*) versus the distance of a user from the BS in meters with $W/R = 10^3$.

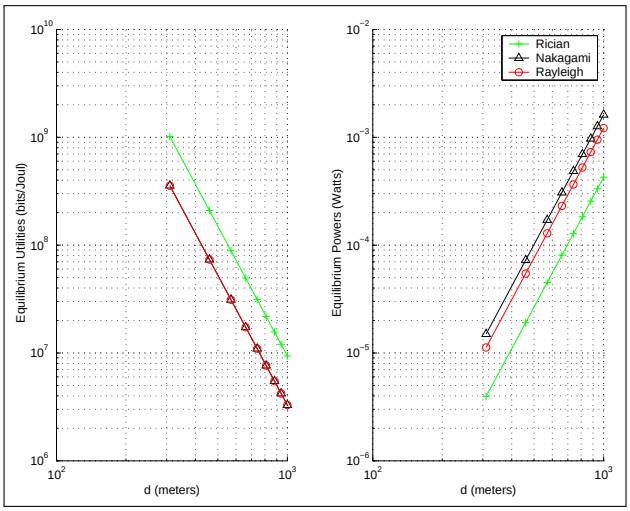


Figure 3: Equilibrium utilities and equilibrium powers of NPG for Rician slow flat fading channel gain (+), and Rayleigh slow flat fading channel gain (o), Nakagami slow flat fading ($\triangle$) versus the distance of a user from the BS in meters with $W/R = 10^3$.

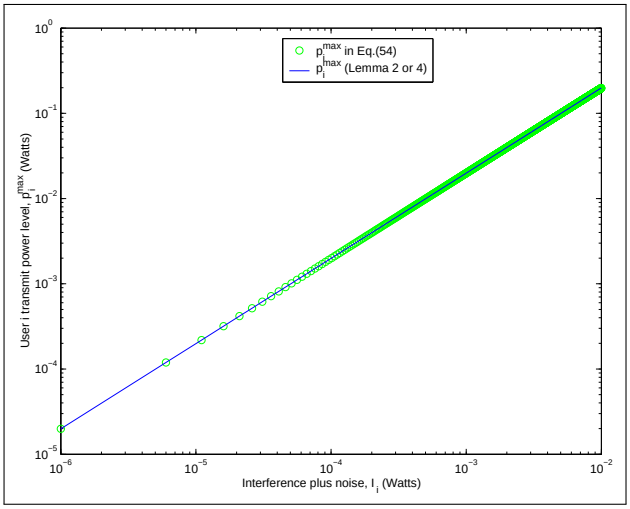


Figure 4: p_i^{max} as a function of I_i as in Eq. (44) (o), and the linear expression $p_i^{max} = 2\,\beta\,I_i$ given in Lemmas 2 and 4 (solid line) with $c = c_{Best}$.

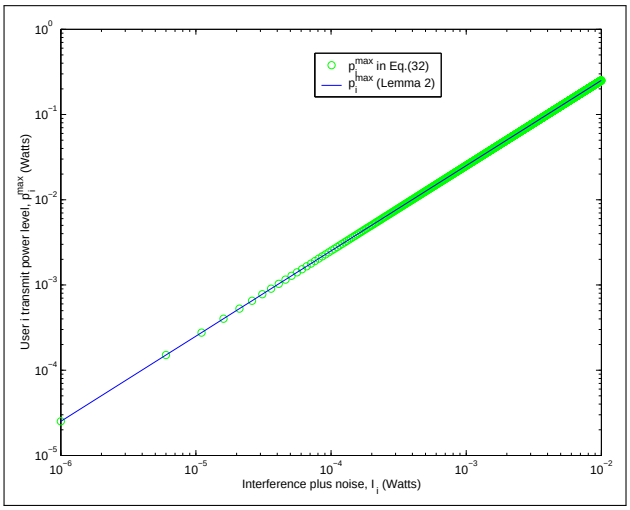


Figure 5: p_i^{max} as a function of I_i as in Eq. (47) (o), and the expression of $p_i^{max} = \sqrt{3\,\xi}\,I_i$ given in Lemma 6 (solid line) with $c = c_{Best}$.

8 Conclusions

We studied a noncooperative power control game (NPG) and noncooperative power control game with pricing (NPGP) introduced in [1] for realistic channel models, where we studied the impact of power statistical variation in Rayleigh, Rician and Nakagami fast/slow flat fading channels on the powers and utilities vectors at equilibrium. The results showed that an equilibrium with an equal maximizing SINR is not attainable in both games with spreading gain ($W/R = 10^2$). In fast fading with spreading gain $W/R = 10^3$, fixed target SINR NPG admitted NE point only under both Rician and Nakagami small scale models. And unlike fast fading, NPG admitted NE point under all small scale fading models in slow fading channels. Results demonstrated that in slow flat fading channels, NPGP with linear pricing does not exhibit a Pareto dominance over NPG outcomes at equilibrium.

In order to overcome the fading effects the SINRs targeted at equilibrium are higher for all users at equilibrium in the Rician, Rayleigh and Nakagami flat fading cases than SINRs under deterministic (nonfading) channels.

Bibliography

[1] C. U. Saraydar, N. B. Mandayam, D. J. Goodman (2002); Efficient power control via pricing in wireless data networks, *IEEE Tras. Comm.*, 50(2): 91-303.

[2] V. Shah, N. B. Mandayam, and D. J. Goodman (1998); Power control for wireless data based on utility and pricing, *Proceedings of PIMRC*, 1427-1432.

[3] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman (2001); Pricing and power control in multicell wireless data network, *IEEE JSAC*, 19(10): 1883-1892.

[4] T. Alpcan, T. Basar,R. Srikant, and E. Altman (2001); CDMA Uplink Power Control as a Noncooperative Game, *Proc. IEEE Conference on Decision and Control*, 197-202.

[5] H. Ji and C.-Y. Huang (1998); Non-cooperative Uplink Power Control in Cellular Radio Systems, *Wireless Networks*, 4(3): 233-240.

[6] M. Xiao, N.Schroff, and E. Chong (2001); Utility Based Power Control in Cellular Radio Systems, *Proc. of Infocomm*, DOI:10.1109/INFCOM.2001.916724, 1: 412-421.

[7] Sarma Gunturi, Paganini, F. (2003); Game theoretic approach to power control in cellular CDMA, *IEEE 58th Vehicular Technology Conference, 2003 (VTC 2003-Fall)*, 4: 2362-2366.

[8] M. Hayajneh and C. T. Abdallah (2004); Distributed Joint Rate and Power Control Game-Theoretic Algorithms for Wireless Data, *IEEE Comm. Letters*, 8(8): 511-513.

[9] C.W. Sung, K.K. Leung (2005); A generalized framework for distributed power control in wireless networks, *IEEE Trans. Info. Theory*, 51(7): 2625-2635.

[10] D. Fudenberg and J. Tirole (1991); *Game Theory*, The MIT Press, 1991.

[11] J. Zander (1992); Distributed cochannel interference control in cellular radio systems, *IEEE Tran. Veh. Tchnol.*, 41: 305-311.

[12] R. D. Yates (1995); A framework for uplink power control in cellular radio systems, *IEEE Journal on Selected Areas in Communication*, 13(7): 1341-1347.

[13] J. G. Proaki (2000); *Digital Communications*, The McGraw Hill Press 1221 Avenue of the Americas, New York, NY 10020, 2000.

[14] Wang, J. T. (2005); Centralized and distributed power control algorithms for multimedia CDMA networks, *International Journal of Communication Systems*, doi: 10.1002/dac.696, 18: 179-189,

[15] T. Alpcan, X. Fan, T. Basar, M. Arcak, J. T. Wen (2008); Power control for multicell CDMA wireless networks: A team optimization approach, *Wireless Networks*, 14(5): 647-657.

[16] Wang, J. T. (2011); Joint rate regulation and power control for cochannel interference limited wireless networks, *International Journal of Communication Systems*, doi: 10.1002/dac.1203, 24(8): 967-977.

[17] Roger L. Peterson, Rodger E. Ziemer and David E. Borth (1995); *Introduction to Spread Spectrum Communications*, Prentice Hall, Upper Saddle River, NJ , 1995.

[18] J. V. Neumann and O. Morgenstern (1994); *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1944.

[19] Don Ross (1999); *What People Want: The concept of utility from Bentham to game theory*, University of Cape Town Press, South Africa, 1999.

[20] I. S. Gradshteyn and I. M. Ryzhik (1980); *Table of Integrals, Series, And Products*, Academic Press, New York 1980.

[21] V. V. Petrov (1995); *Limit Theorems of Probability Theory: Sequences of Independent random variables,* Clarendon Press, Oxford, 1995.

[22] https://www.dropbox.com/s/62uyd038cjnrwss/Hayajneh_supplementary.pdf.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):539-550, August, 2015.

Understanding Social Characteristic from Spatial Proximity in Mobile Social Network

D. Hu, B. Huang, L. Tu, S. Chen

Duan Hu,Benxiong Huang,
Lai Tu, Shu Chen*
Department of Electronics and Information Engineering
Huazhong University of Science and Technology
1037 Luoyu Road, Wuhan, China
huduan1985@gmail.com, huangbx@hust.edu.cn,
laitu@hust.edu.cn, remark.chen@gmail.com
*Corresponding author: remark.chen@gmail.com

Abstract: Over the past decades, cities as gathering places of millions of people rapidly evolved in all aspects of population, society, and environments. As one recent trend, location-based social networking applications on mobile devices are becoming increasingly popular. Such mobile devices also become data repositories of massive human activities. Compared with sensing applications in traditional sensor network, Social sensing application in mobile social network, as in which all individuals are regarded as numerous sensors, would result in the fusion of mobile, social and sensor data. In particular, it has been observed that the fusion of these data can be a very powerful tool for series mining purposes. A clear knowledge about the interaction between individual mobility and social networks is essential for improving the existing individual activity model in this paper. We first propose a new measurement called geographic community for clustering spatial proximity in mobile social networks. A novel approach for detecting these geographic communities in mobile social networks has been proposed. Through developing a spatial proximity matrix, an improved symmetric nonnegative matrix factorization method (SNMF) is used to detect geographic communities in mobile social networks. By a real dataset containing thousands of mobile phone users in a provincial capital of China, the correlation between geographic community and common social properties of users have been tested. While exploring shared individual movement patterns, we propose a hybrid approach that utilizes spatial proximity and social proximity of individuals for mining network structure in mobile social networks. Several experimental results have been shown to verify the feasibility of this proposed hybrid approach based on the MIT dataset.
Keywords: Mobile social network; Geographic community; Community structure; Measurement;

1 Introduction

Recent years, with the increasing popularity of mobile terminal devices , the emerging massive urban sensors such as GPS, mobile phone and online user-generated social media enable us understand the intrinsic nature of human mobility such as high predictability in daily routine and the social demographic characteristics of the individuals. Given the importance of human mobility, e.g. human movement could be responsible for the geographical spread of human infectious diseases [2], many researchers have focused on capturing the statistical characteristics of individual human trajectories [3, 4], understanding the role of geographic distance and social interaction in human spatial redistribution on geographic scales[5-7]. Early empirical studies on albatrosses, monkeys and marine predators [8] suggested that animal trajectory can be approximated by a levy fight [9]. Given that money is carried by individuals, Brockmann [1] et al. illustrated individual geographic trajectory using half a million dollar bills, suggesting that

human trajectories could be best modeled as a continuous-time random walk (CTRW model) with fat-tailed displacements and waiting-time distributions.

However, several types of empirical data [10]on human mobility, captured by mobile phone traces, show that the predictions of the CTRW models are systematically in conflict with the empirical results. Due to the fundamental importance of statistical properties of human travel, in recent years, the increasing pervasiveness of location-acquisition technologies such as 3G network and GPS is conducive to the collection of large spatial-temporal datasets and opening up new opportunities for discovering some different knowledge about movement behaviors. Indeed, the availability of these collected large-scale call data such as mobile phone records and global-positioning-system data, has offered researchers from various disciplines access to detailed patterns of human behavior, and greatly enhanced our understanding of human mobility [3]. Recently, distinguished from previous studies, a new understanding [4] about building individual predictive model for future locations suggests that at a global scale human mobility exhibits structural patterns subject to geographic and social constraints.

Researchers find that humans experience a combination of periodic movements limited by geographical factors and seemingly random jumps correlated with their social networks [4]. In this case, individual movements can be divided to the two parts: short-range travel driven by periodic movements and long-range travel driven by social interaction. Meanwhile, some studies [4, 5] found that individual (mobile devices) mobility patterns can shape and impact the inherent properties of social network. Similarity of individual mobility patterns, which is called "spatial proximity", is expected to exert a significant impact on the inner structure of such social networks, from the maintenance of long-lasting friendships to the formation of new links. It seems that social links are often driven by spatial proximity, from job- and family-imposed programs to joint involvement in various social activities [11]. On the contrary, the degree of similarity in periodic movements and social proximity of individuals can positively affect the level of spatial proximity since the similarity of both types of individual movements can be identified by spatial proximity. Based on collecting large-scale call data that are recorded, spatial proximity between any two mobile phone users has been used to predict the potential call in cell phone social networks [5].

The booming development of mobile social network gives rise to several amazing sub-issues, such as community detecting, link prediction, social-aware recommendation. This was explored through evaluating social property of individuals and mining network structure evolution using general social network analysis. However, given those fascinating features of spatial proximity for forming social interaction among individuals in mobile social network that has been revealed, it is just natural to consider the social influence of spatial proximity that can be used to improve current researches on those amazing sub-issues.

Therefore, the aim of this paper is to investigate the social features of spatial proximity and further mine social network structure evolution to explore some potential possibility for optimizing user experience of mobile social networking service and social recommendation in mobile social network. Although our present work is only confined to analyzing and applying the significant correlation between spatial proximity and social properties of individuals in mobile social network by several well-known data mining methods based on two real datasets, our experimental results also deepen our understanding about the social influence of spatial proximity in mobile social network.

The contributions of this paper are as follows:

- We propose a new measurement dimension called geographic community for clustering spatial proximity in mobile social network. A novel approach for detecting these geographic communities in mobile social network has been proposed. Through developing a spatial proximity matrix, an improved symmetric nonnegative matrix factorization method (SNMF)[13] is used to detect geographic communities in mobile social network.

- We suppose that geographic community can provide a bridge between spatial proximity and social properties of individuals in mobile social network. By a real dataset containing thousands of mobile phone users in a provincial capital of China, the correlation between geographic community and common social properties of users has been shown.

- By mining shared individual movement patterns, we propose a hybrid approach that utilizes spatial proximity and social proximity of individuals for mining network structure in mobile social network. We first propose two new measurement matrixes. Combing these matrixes with the known adjacency matrix in social network, the joint nonnegative matrix factorization (JNMF) method has been used to mine network structure evolution. Finally, a collaborative matrix has been obtained to analyze the existing community structure and potential links in mobile social network.

The rest of this paper is organized as follows: section 2 describes a well-known index of spatial proximity. The two measurement matrixes for evaluating spatial proximity and social ties in such mobile social networks will also be formed. Section 3 presents the definition of our new measurement dimension "geographic community", and the strong correlation between geographic community and the similarity of several key indexes from the socio-economic background of those individuals based on thousands of cell phone call records and corresponding user information. Section 4 first provides a hybrid approach to mine network structure in mobile social network, and then show several experimental results to verify the feasibility of this proposed hybrid approach based on the MIT dataset. Section 5 provides related work and finally comes to the conclusion.

2 Preliminaries

In this section, we first choose spatial cosine similarity to represent the level of spatial proximity of any two individuals. The measurement matrixes that describe such spatial proximity of corresponding individuals have been presented. Furthermore, measurement matrixes that utilize the relationship between locations and users for evaluating spatial proximity and social ties in such mobile social networks have also been formed.

Spatial cosine similarity[5]:SCos(x,y)

The cosine similarity of user x and y's trajectories, describing how similar their visitation frequencies are, is represented by the cosine of the angle between the two vectors of number of visits at each location for x and y and,

$$SCos(x,y) = \sum_{l=Loc} \frac{PV(x,l) \times PV(y,l)}{||PV(x,l)|| \times ||PV(y,l)||}. \tag{1}$$

Where $PV(x,l) = \sum_{i=1}^{n(x)} \delta(l, L_i(x))/n(x)$ is the probability that user x visits location l, Loc is the set of all locations (cell phone towers and etc.).

Spatial proximity matrix

In order to reflect the spatial proximity between any two individuals in mobile social networks, we propose an approximate overlay $G = (V, E, w)$, then the adjacency matrix of G can be written as:

$$G = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix} \tag{2}$$

where $s_{ij}$is spatial cosine similarity $SCos(i,j)$, i ,j is each individual respectively. Specially, when $i = j$,$SCos(i,j) = 1$. Note that $SCos(i,j) = SCos(j,i)$, G is a symmetric nonnegative matrix.

The two measurement matrixes: user-location matrix and location-user matrix

In previous studies, for evaluating the similarity in mobility patterns of two individuals, several indexes of spatial proximity have been proposed by the references [5, 14], spatial proximity can be defined by one key index of them:

The cosine similarity of user x and y's and s trajectories, describing how similar their visitation frequencies are, is represented by the cosine of the angle between the two vectors of number of visits at each location for x and y:

$$SCos(x,y) = \sum_{l=Loc} \frac{PV(x,l) \times PV(y,l)}{||PV(x,l)|| \times ||PV(y,l)||}. \tag{3}$$

Where $PV(x,l) = \sum_{i=1}^{n(x)} \delta(l, L_i(x))/n(x)$ is the probability that user x visits location l, Loc is the set of all locations (cell phone towers and etc.).

As mentioned above, since the internal relationship between individuals and their usual travel locations can form spatial proximity, which has a significant impact on network structure of mobile social network, we develop two measurement matrixes that can reflect such internal relationship to evaluate network structure. Given N individuals (or mobile personal devices) in a mobile social network and the number of all locations, we propose a user-location matrix U. (User-location matrix)

$$U = \begin{pmatrix} PV_{11} & \cdots & PV_{1M} \\ \vdots & \ddots & \vdots \\ PV_{N1} & \cdots & PV_{NM} \end{pmatrix} \tag{4}$$

where $PV_{xl} = PV(x,l), x \leq N, l \leq M$

(Location-user matrix)

$$L = \begin{pmatrix} LV_{11} & \cdots & LV_{1N} \\ \vdots & \ddots & \vdots \\ LV_{M1} & \cdots & LV_{MN} \end{pmatrix} \tag{5}$$

Where $LV_{xl} = \sum_{i=1}^{n(l)} \delta(l, L_i(x))/n(x), x \leq N, l \leq M$ is the probability that user x visits location l, Loc is the set of all locations (cell phone towers and etc.).

3 Geographic community and its correlation with the social properties of individuals

In this section, we first propose a new measurement dimension for clustering those individuals with close spatial proximity in mobile social network. Based on SNMF method, an approach for detecting these geographic communities has been proposed. Several numerical experiments have been shown to verify the feasibility of this measurement approach.

Geographic community

Given N individuals (mobile personal devices) in a mobile social network, the similar degree of geographic trajectories of two individuals can be measured by the spatial proximity index. In this case, we can develop an approximate overlay which indicates the spatial proximity between two individuals (nodes) in the mobile social network.

Suppose $G = (V, E, w)$ is an undirected weighted graph with N nodes and M links representing this approximate overlay. It is quite different from general social graphs in that these links between any two nodes are the spatial proximity index. The weight $w \in [0, 1]$ of the edge E denotes the similar degree of two individuals which can be calculated by spatial cosine similarity.

$C = C_1, C_2, \ldots, C_k$ denotes a collection of disjoint geographic communities, where $C_i \in C$ is a geographic community of G.

Using general definition of community structure in social network, the geographic community has several meanings: the nodes in geographic community are those individuals which have higher similar degree, that is, those individuals had met each other frequently in the past long period of time, unlike those individuals between different geographic communities. The strong correlation between this geographic similarity and their proximity in social network indicates that more social connections among the nodes will occur in geographic community in the future. Therefore, geographic community can be a measurement dimension that estimates the interaction between social community and spatial proximity. In fact, we consider geographic community as an indicator which may be used for evaluating the evolution of social community structure in the future.

Detecting geographic community using symmetric nonnegative matrix factorization (SNMF)

We consider symmetric NMF as our tool to find the geographic community because of its powerful interpretability and close relationship with clustering methods. It has been verified that SNMF has outstanding performance in detecting communities in undirected unweight complex networks.

Given that our approximate overlay is an undirected weighted virtual network, we further apply SNMF method to achieve our purpose. If G is symmetric, then S is symmetric. We can then absorb S into X. i.e. $\hat{X} = XS^{1/2}$. Then our task is to solve the following problem:

$$\min_{x \geq 0} \|G - \hat{X}\hat{X}^T\|_F^2 \tag{6}$$

According to [15],$\hat{X}$ can be obtained by the following multiplicative update rule

$$\hat{X}_{ik} \leftarrow \hat{X}_{ik} \left[\frac{1}{2} + \frac{(G\hat{X})_{ik}}{(2\hat{X}\hat{X}^T\hat{X})_{ik}} \right] \tag{7}$$

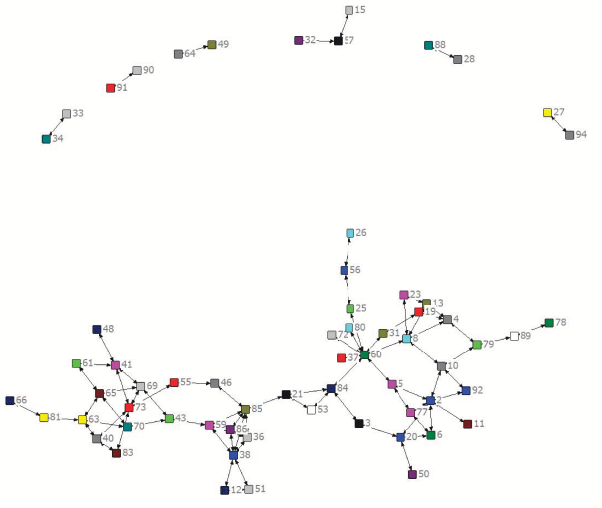


Figure 1: The reported friendship network in Reality Mining datasets

After iteration, the obtained $\hat{X}^*$ is just the scale partition matrix of the network G of size $n \times K$, whose i-th row corresponds to the cluster (geographic community) membership of the i-th unit.

We can further normalize $\hat{X}$ to make $\sum_j \hat{X}_{ij} = 1$ so that $\hat{X}_{ij}$ corresponds to the posterior probability that the i-th unit belongs to the k-th geographic community.

We choose Reality Mining data set [6] provided by the MIT Media Lab to test our proposed approach. The Reality Mining dataset contains communication, proximity, location, and activity information from 94 students or teachers at MIT during the 2004-2005 academic years. In particular, the dataset includes call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and so on. In this dataset, a mobile social network which contains 94 nodes and several million links can be modeled as Fig.1, where the links representing real social friendship which can be evaluated by the number of calls, messages and so on.

To design this numerical experiment, we propose those cell tower ID lists of each individual communication record representing those locations l that can be used to calculate the probability $PV(x, l)$.We firstly obtain the spatial proximity matrix of this real mobile social network, then an approximate overlay network can be inferred in Fig.2. As is shown in Fig.2, this network does not have obvious community structure. We consider the reasons as follows: every two individuals have their spatial similarity. In the overlay network, the weights of links are inferred by corresponding spatial proximity. In this case, it is essential to properly evaluate the impact of the weight value on network structure. The binarization method will lose much useful information. Furthermore, these individuals have more overlaps in geographic community than general communities. Therefore, we choose to delete those weight links which have lower weights (i.e., the links that weight $w \leq 0.2$ will be deleted).

We use SNMF method for detecting geographic community on this overlay, the symmetric nonnegative matrix G is a $n \times n(n = 94)$ spatial proximity matrix. We assume the number of main geographic communities $K = 2$, the $K \times n$ matrix will be calculated by (7). The final iteration result of X is shown in Fig. 2.

In Fig. 3, the interval $[0, 1]$ located in the Y-axis represents the community $ID = 1$, the interval [1,2] located in the Y-axis represents the community $ID = 2$, the sequence of nodes $i(i = 1, 2, \ldots, 94)$ is located in the X-axis, and the gray level is obtained by corresponding element value of X , where the higher the gray level of the block $(i, ID(i))$tend to be, the more

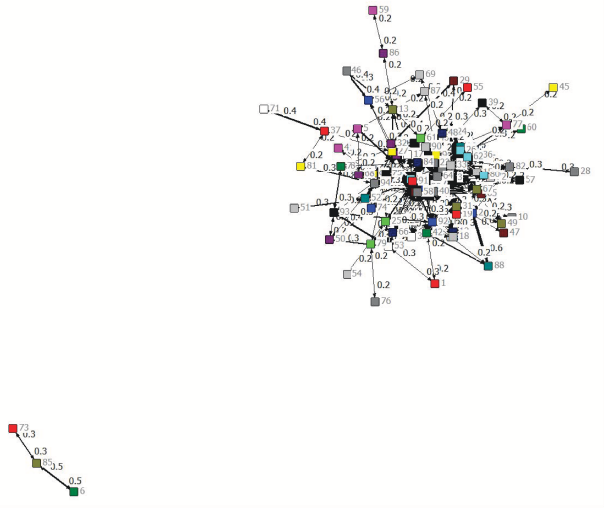


Figure 2: the spatial proximity overlay of reality mining datasets

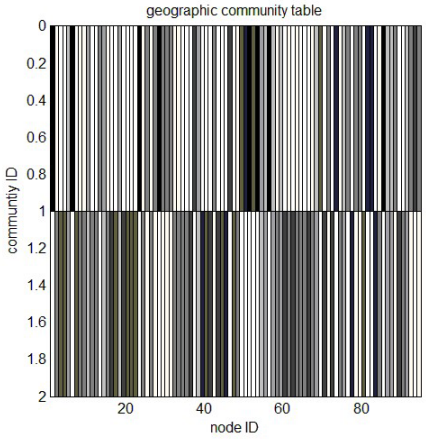


Figure 3: the results obtained by applying our proposed method

likely that unit i belongs to the geographic community ID.

The relevance test between geographic community and the social properties of individuals

We use a dataset that captures for two-month period the time-resolved trajectories of ten thousand anonymized mobile phone users in an anonymous provincial capital in China to uncover the relevance between geographic community and the social properties of individuals.

Given the multiplicity of the social properties of individuals and this dataset mainly contains the cell phone charging records of all users in an anonymous telecommunications company, the similarity of the cell phone charging of these users depends on the similarity of some social properties such as the economic ladder to some extent. Therefore, we choose the type of plan the user used and the interval of user charging per month as the two parameters of the economic ladder and present the value of the economic ladder of them. Three geographic communities have been obtained from ten thousand users by the above approach based on this dataset.

Through clustering these economic ladder values in each geographic community, we found that the economic ladder of individuals is highly consistent in each geographic community.

In this test, our parameters are set as follows:

Table 1: The initial situation of ten thousand mobile phone

The number of individuals	Economic ladder	Matching results in general
10000	$0(budget)(ARPU < 100)$	5613(56.13%)
	$1(ordinary)(100 < ARPU < 200)$	3052(30.52%)
	$2(luxury)(ARPU > 200)$	1335(13.35%)

Finally, our test truly shows the higher consistency in table2, given the number of geographic communities $N_c = 3$, and we consider the posterior probability of nodes $\hat{X}_{iC} > 0.4$, $C \in 0, 1, 2$ means node i belong to C,then we obtained following numerical results:

Table 2: The matching results between the geographic communities and economic ladder

The label of geographic communities	Matching results		
$0(number : 5767)$	63.4%	28.85%	7.6%
$1(number : 4978)$	56.3%	40.1%	3.57%
$2(number : 3244)$	44.1%	35.2%	20.6%

Given that periodic movements can explain about 50-70% of the geographic trajectories of individuals, and the repetitive trajectories that are formed by periodic movement mainly depend on the role of daily routine and geographic features of human movement, we consider the similarity of the geographic trajectories between individuals is largely attributable to the overlapping functional areas for these individuals, for example, identical residential area, the same lunch venue and similar workplace. Furthermore, these overlapping functional areas just depend on the social properties of individuals, such as economic ladder, vocation, and so on. Therefore, the spatial proximity of individuals not only suggests the proximity of social relationship of individuals in mobile social network, but also implies to some extent the similarity of the social properties of individuals. Furthermore, we suggest that the similarity of the social properties of users also indicates the likeliness for these users to interact with each other.

4 A hybrid approach to mine network structure in mobile social network

Since SNA techniques have gained extensive development in recent years, many studies on detecting existing community structure in various social networks can be found in an excellent survey [16]. Furthermore, a large amount of studies on predicting potential links have been presented for link prediction in complex networks [17]. Mining network structure, which provides us with meaningful insights into the internal structure of corresponding real social network and the possibilities to optimize the performance of social recommendation and link prediction, mainly contains two parts, i.e. detecting existing network structure and tracking the trend of network structure evolution. In general, mining network structure in mobile social networks has been investigated as a typical example of dynamic social networks by many researchers [6, 18].

Spatial proximity and social proximity of individuals can be used for mining network structure in mobile social network. Especially, existing community structure and potential social relationship in a real mobile social network can be obtained and analyzed by our hybrid approach. We first propose two new measurement matrixes which provide a bridge between spatial proximity and the social properties of individuals in mobile social networks. Combing these matrixes and the known adjacency matrix in social network, the joint nonnegative matrix factorization (JNMF) method [13] is used to mine network structure evolution. Finally, a collaborative matrix is obtained to analyze the existing community structure and potential links in mobile social network. Based on a real dataset, several experimental results have been shown to verify the feasibility of this proposed hybrid approach.

Mining community structure in mobile social network using joint nonnegative matrix factorization (JNMF) method

In previous studies, JNMF was usually used to detect community structure or hidden links in complex networks. In fact, it has been verified that JNMF has outstanding performance in such aspects in complex networks. In this section, we choose JNMF method as our tool to find community. Based on the MIT dataset, this social graph is an undirected weight network. Given its adjacency matrix of friend graph D, we further apply JNMF method to achieve our purpose. The problem we aim to solve is

$$\min_{x \geq 0} l(X, U, D, L)$$
$$\text{s.t. } X \in \mathbb{R}_{+}^{n \times m}$$

Where $l(X,U,D,L)^{def} = \|U - X\|^2 + \alpha\|D - XX^T\|^2 + \beta\|L - X^TX\|^2$, and $\alpha > 0$,$\beta > 0$ are constants to tradeoff the importance between different terms. Given N individuals (or mobile personal devices) in a mobile social network, and the number of all locations m , and D represents the adjacency matrix of friend graph , U and L represent User-location matrix (4)and location-user matrix (5), respectively . According to [15], $\hat{X}$ can be solved by the following multiplicative update rule.

$$\hat{X}_{ij} \leftarrow \hat{X}_{ij} \left[\frac{[U + 2\alpha DX + X\hat{L}]_{ij}}{2(\alpha + \beta)[XX^TX]_{ij}} \right]^{\frac{1}{4}} \tag{8}$$

where $\hat{D} = 2\beta D - 1$

After iteration, the obtained $\hat{X}^*$ is just the scale partition matrix of the network G of size $N \times M$, and then we obtain the hybrid matrix $C = X^TX$. Let this matrix is the initial adjacent matrix, we further use the above approach in section 3 for detecting community structure based on SNMF.

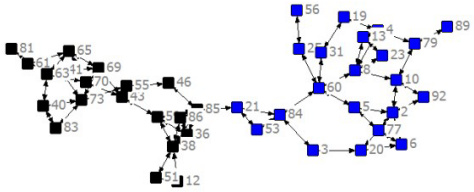


Figure 4: Two main communities in the community structure of this friendship network

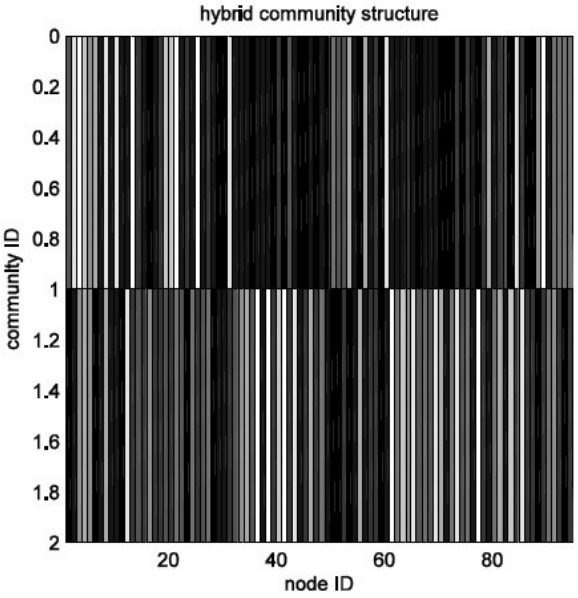


Figure 5: The hybrid communities structure in this MIT dataset

In our experiment, we also use Reality Mining data set [6] provided by the MIT Media Lab to test our proposed approach. From this real friendship network, two main communities can be inferred by real social connections as shown in Fig.2, which shows the first-year business school students and individuals working together in the same building respectively. Moreover, we can simply detect these social communities by Newman method. However, the dynamic evolution of the social ties among these individuals cannot be timely detected. This leaves an important question that to what extent individual mobility patterns impacts the social network. Therefore, in this section, the communities of the same dataset will be detected by our proposed method.

During our implementation process, we suppose spatial proximity effect on social community structure is limited, and present two weight parameters. As is shown in Fig.5, the difference in gray-level between two communities labels for nodes illustrates the segmentation map presented in Fig.4. Meanwhile, some nodes (ex:$i = 2, 33, 76$) that are not covered by this segmentation map shows the potential possibility for joining the two communities together.

In Fig. 5, the interval $[0, 1]$ located in the Y-axis represents the community $ID = 1$,and the interval $[1, 2]$ located in the Y-axis represents the community $ID = 2$, the sequence of nodes $i(i = 1, 294)$ is located in the X-axis, and the gray level is obtained by corresponding element

value of X , where the higher the gray level of the block $(i, ID(i))$ tend to black, the more likely that the unit i belongs to the geographic community ID .

5 Results and Discussion

Since recent research progress on the interaction between human mobility and social ties has provided us with valuable information that spatial proximity can be used to predict future social links, in this paper, we first proposed a new measurement dimension (geographic community) for evaluating the spatial proximity influence on the evolution of social community structure in mobile social networks.

The correlation between geographic community and common social properties of users has been analyzed based on a real dataset. We found the spatial proximity of individuals not only suggests the proximity of social relationship of individuals in mobile social networks [5], but also implies to some extent the similarity of the social properties of individuals. Then a hybrid approach that utilizes spatial proximity and social homophily of individuals for mining network structure in mobile social network has been proposed.

However, due to the inner ambiguity of individuals spatial proximity, the direct applicability of spatial proximity in mobile social network still leaves much room to be studied. The results presented in this paper will lead us to two interesting directions for future research. The first direction is to explore the common social properties of individuals based on those common functional areas that can be obtained from spatial proximity, since we consider high spatial proximity should be driven by similar social feature of individuals to a certain extent and even develop a hybrid technique for personalized recommendation utilizing geographic locations and social network. The second direction is to develop a routing algorithm on MANET (mobile ad-hoc network) using the inherent property of geographic community, since this communication services that rely on this type of data transfer will strongly depend on human mobility characteristics and how often such transfer arises.

Bibliography

[1] D. Brockmann, L. Hufnagel, and T. Geisel (2006); The scaling laws of human travel, *Nature*, 439:462-465.

[2] L. Hufnagel, D. Brockmann, and T. Geisel (2004); Forecast and control of epidemics in a globalized world, *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124-15129.

[3] C. Song, Z. Qu, N. Blumm, A. Barabasi (2010); Limits of predictability in human mobility, *Science*, 327(5968): 1018-1021.

[4] E. Cho, S. A. Myers, and S. J. Leskovec(2011); Friendship and mobility: User movement in location-based social networks, *Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, USA, 1082-1090.

[5] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A. Barabasi (2011); Human mobility, social ties, and link prediction, *Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, USA, 1100-1108.

[6] N. Eagle, A. Pentland, and D. Lazer (2009);Inferring friendship network structure by using mobile phone data, *Proc. of the National Academy of Sciences*, 106(36): 15274-15278.

[7] L. Backstrom, E. Sun, and C. Marlow (2010); Find me if you can: improving geographical prediction with social and spatial proximity, *Proc. of the 19th international conference on World wide web(WWW'10)*, New York, USA, 61-70.

[8] M. C. Gonzalez, C. A.Hidalgo, and A. Barabasi (2008); Understanding individual human mobility patterns, *Nature*, 453: 779-782.

[9] R. N. Mantegna and H. E. Stanley (1994); Stochastic process with ultraslow convergence to a Gaussian: the truncated Levy flight, *Physical Review Letters*, 73: 2946-2949.

[10] C. Song, T. Koren, and A. Barabasi (2010); Modelling the scaling properties of human mobility, *Nature Physics*, 6: 818-823.

[11] M. T. Rivera, S. B. Soderstrom, and B. Uzzi (2010); Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms, *Annual Review of Sociology*, 36: 91-115.

[12] Q. Hao, et al. (2010); Equip tourists with knowledge mined from travelogues, *Proceedings of the 19th international conference on World wide web (WWW'10)*, New York, USA , 401-410.

[13] F. Wang, et al. (2011); Community discovery using nonnegative matrix factorization, *Data Mining and Knowledge Discovery*, 22: 493-521.

[14] Q. Li, et al. (2008); Mining user similarity based on location history, *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems(GIS'08)*, Irvine, CA, USA, 34-43.

[15] D. Wang, T. Li, S. Zhu, and C. Ding (2008); Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization, *Proc. of the 31st annual international ACM SIGIR conference on Research and development in information retrieval(SIGIR'08)*, New York, NY, 307-314.

[16] S. Fortunato (2010); Community detection in graphs, *Physics Reports*, 486: 75-174.

[17] L. Lu and T. Zhou (2011); Link prediction in complex networks: A survey, *Physica A: Statistical Mechanics and its Applications*, 390: 1150-1170.

[18] N. P. Nguyen, et al. (2011); Adaptive algorithms for detecting community structure in dynamic social networks, *Proc.IEEE INFOCOM*, Shanghai,China, 2282-2290.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):551-566, August, 2015.

Analytical Modelling and Performability Evaluation of Multi-Channel WLANs with Global Failures

Y. Kirsal-Ever, Y. Kirsal, E. Ever, O. Gemikonakli

Yoney Kirsal-Ever, Orhan Gemikonakli
Computer and Communication Engineering
Middlesex University
London, Hendon, UK
y.kirsal@mdx.ac.uk,o.gemikonakli@mdx.ac.uk

Yonal Kirsal
Electrical and Electronic Engineering
European University of Lefke
North Cyprus, Mersin 10, Turkey
ykirsal@eul.edu.tr

Enver Ever*
Computer Engineering
Middle East Technical University
Northern Cyprus Campus, Guzelyurt, Mersin 10, Turkey
*Corresponding author: eever@metu.edu.tr

Abstract: Wireless local area networks (WLANs) which are based on IEEE 802.11 standard are used widely in existing local area network configurations. IEEE 802.11 offers multiple non-overlapping channels to increase the capacity of the network. There are strong evidences that WLANs are prone to impairments. In order to improve the quality of service (QoS) and to evaluate the performance of WLANs realistically, the availability of the systems should be considered. This paper studies performability evaluation of a multi-channel WLAN using analytical modelling approach. Unlike the existing studies, the failures of the overall system, where a critical function unit fails making all the channels unavailable are considered. A new term is introduced as global failures. It is possible to solve the models considered using matrix geometric method where system parameters and minimal non negative solution R is computed by an iterative method. However spectral expansion method is a well-known alternative where the iterative calculations for solving R is avoided using eigenvalues and eigenvectors. The exact spectral expansion method is employed to obtain performability measures such as mean queue length and blocking probability. Iterative refinements are employed in solution of simultaneous equations.
Keywords: Analytical modelling, 2-dimensional Markov chain analysis, performability evaluation, availability, spectral expansion, multi-channel WLANs.

1 Introduction

The demand of new multimedia services, increase in bandwidth available to users, high traffic densities, and ubiquitous connectivity have caused the rapid growth of wireless and mobile communication systems. Wireless LANs (WLANs) based on the IEEE 802.11 standard have gained widespread popularity mainly because they provide users unlimited access to the Internet with relatively high data rates. The IEEE 802.11 specifications provide a multi-rate capability at the physical layer (PHY) to accommodate mobile users with diverse wireless channel conditions [1]. The IEEE 802.11b specification supports data rate up to 11 Mbps. The subsequent versions of WLANs such as 802.11a, 802.11g, 802.11e and the latest version, 802.11n, offers maximum

rate of 54 Mbps and 500-600 Mbps respectively [2, 3]. The classical IEEE 802.11 MAC protocol considers a single shared channel. Nevertheless, multiple channels can also be used [1,5,20,22,24]. The recent WLANs based on IEEE 802.11 standards permit simultaneous operation on multiple non-overlapping channels at the physical layer. Extensive research has been carried out on multi-channel MAC protocols for IEEE 802.11 based WLANs [1, 21, 23, 24]. These approaches proposed multi-channel WLANs to improve overall network performance depending on different channel assignment schemes. In other words, the wireless access points (APs) can be considered as multi-channel systems with multiple wireless radios each operating independently on different channels. These multiple radios and multiple channels facilitates dynamic selection of channels for the users to transmit and receive packets without interfering each other [1, 2, 4, 5, 8, 14]. Thus, these new features improve the QoS such as throughput, blocking probability, response time etc. [15, 21–24].

In [20] two cross-layer models have been proposed to access transmissions channels. The first model is single class type traffic. The traffic divided into two classes as high and low priority traffic in the second model. The performance parameters for both single and multi-channel wireless networks, like the requests throughput, acceptance probability are calculated. WLANs suffer especially when the users with different and high transmission rates share the same physical channel. In order to solve this problem different transmission channels have been proposed in [21] to achieve better QoS. In addition, an 802.11 based multi-hop wireless ad-hoc network architecture have been proposed in [22] where it employs multiple radio channels simultaneously in the system. Multiple channels along with the bi-directional channel reservation policy of ad hoc networks based on IEEE MAC protocol considered is considered in [23] using analytical modelling. The paper also considered two different channel schemes for further analysis. In [24] existing multi-channel protocols are proposed to enhance the QoS by classifying them into single-radio protocols, multi-radio single-hop protocols, and multi-radio multi-hop protocols. The existing studies in the literature propose modified MAC layer architecture which supports multi-channel networks. The main aim is to find an optimal channel for a single packet transmission to improve QoS.

In order to improve QoS and to maximize the performance of such systems analytical modelling and evaluation is still one of the key issues in the performance characterization of wireless communication systems. Various approaches exist in the literature for the performance evaluation of WLANs [6–8, 12, 14, 20]. However, most of the existing work has been done on pure performance analysis of IEEE 802.11 distributed coordination function (DCF) as well as the analysis of the IEEE 802.11 MAC protocol. The traditional pure performance models are too optimistic. They ignore failure and recovery behaviour and only focus on system's ability to perform. System availability models, that consider failure and repair, should be taken into account in order to obtain more realistic QoS measures. The performability modelling and evaluation is a combination of performance and availability models [9–11, 13]. In practice, WLANs are prone to system failures such as failure of an AP and base station (BS) like any other physical system [10, 11, 13, 15, 16, 19]. The signal strength, fading, reflection and refraction of signals, and man-made noise etc. are some important factors that can be the cause of the channel unavailability in WLANs [5, 10, 11, 13, 15]. In WLANs, failing of AP/BS and/or the components of networks may cause the failure of all the available channels. This fault is termed as global system failure of WLAN in this paper. Examples of system failure scenarios in WLANs would include failure of APs and/or BSs, loss of the link between APs and users etc. This type of failures may occur when an AP fails or the communication links between an AP and service providers do not function, or operate. Failures may occur due to software, hardware, human error, or a combination of these factors in WLANs. Different kind of failures, reasons, and failure affects can be found in [5, 9, 11–13, 16, 19]. Channel failures due to the environmental interference as

well as from other wireless devices are considered in [15]. Moreover, in [13], channels are defined as virtual channels and each channel is subject to failure for transmission because of multipath fading in the WLANs.

In case of a system failure, all the channels, resources, services, and the components become unavailable to the users. System and channel failures occurring in system is lead to degradation of the entire system performance. However, the existing studies do not consider the channel unavailability with global failures which can be quite critical for QoS measurements. One of the main objectives in this paper is to model and analyse the system performance considering global failures and recovery in multi-channel WLANs. Thus, modelling and analysis of the system performance with global failures and repairs is considered with potential channel failures in multi-channel WLANs. In addition, mobility related factors are considered and analysed. Therefore the proposed model attempts to understand, and improve system performance in terms of system capacity, mobility and blocking probability of packets. The channel assignment scheme employed is another major issue in the performance characterisation of wireless communication systems. There have been extensive studies on dynamic channel access in multi-channel WLANs recently [1, 3, 4, 14, 20, 22, 23]. This is because dynamic channel access scheme is appropriate and efficient scheme to improve QoS of multi-channel WLANs. There is no fixed relationship between channels and systems in dynamic channel access. All channels are assigned into a central pool (e.g bandwidth). The channels are assigned to the incoming packet requests dynamically and channel returns to the central pool after the service [3, 4]. In order to meet the future demands and increase available bandwidth for each user within WLANs, dynamic channel access is also taken into consideration in this paper. This is done by (assuming the packets are accepted with certain bandwidth assigned) optimal allocation of WLAN bandwidth for the number of channels required. All the channels used by the same WLAN are perfectly coordinated and synchronised.

Spectral expansion method is employed for the exact solution of the analytical model considered [9,10]. Using the proposed model and an exact solution, important performability measures, such as mean queue length and blocking probability can be computed.

The rest of the paper is organized as follows: Section II describes the proposed model. The two dimensional modelling approach and exact steady state solution are explained in section III and section IV respectively. In section V, numerical results computed by using an exact solution approach are presented. Some pure performance results are also presented for comparison. Conclusions and recommendations are provided in section VI.

2 The System Considered and The Analytical Model

The WLAN presented is subject to failures. The entire system failures (e.g AP or BS) are considered as global failures that may be caused by power cut-off or resetting AP etc. The model covers mobility issues and various queue capacities as well. The system has multiple identical channels. Allocation of packets is usually done by availability of channels and in this regard, it is well known that, in terms of efficiency the common queue is more suitable than individual queues in the queuing theory [7, 10, 11, 17]. The proposed model considered for performability evaluation of multi-channel WLANs is shown in Fig. 1.

The system consists S identical channels, numbered 1, 2, 3, $\cdots$, S with a common queue. The common queue is bounded with a capacity of W ($W \geq S$). The maximum number of calls in the system is equal to the number of calls assigned to the channels plus the queuing capacity. This is given by L where, $L = S+W$. The superposition of all arrival streams (potential arrivals from other wireless technologies can be incorporated) follows a Poisson process. In other words, inter-arrival times of the incoming call requests are assumed to follow an exponential distribution with mean rate of λ similar to studies [7–11], [17]. If the channels are available and idle in the

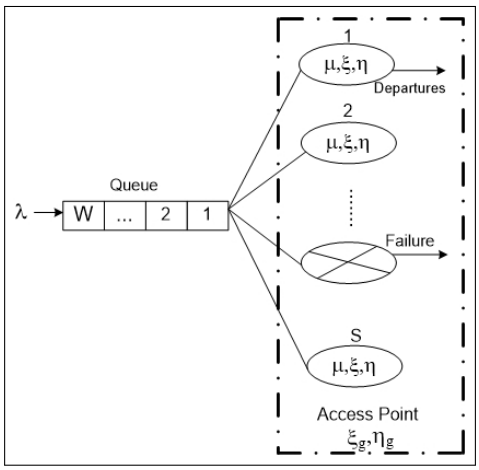


Figure 1: The queueing system considered for proposed multi-channel WLAN

WLAN, the packet arrivals can be assigned to any channel. Otherwise, in case of a channel failure, or a busy channel, the incoming packet request is queued. Similarly, the service times of the packet serviced by channel s (s=1, 2, $\cdots$, S) are distributed exponentially. In addition to that, the dwell time, which is the time that mobile users spend in the WLAN coverage area, is assumed to be distributed exponentially with mean $1/\mu_{wd}$ [9, 10]. Equation (1) is used to calculate μ_{wd}.

$$\mu_{wd} = \frac{E[v]F}{\pi O} \tag{1}$$

Where, $E[v]$ denotes the expected value of the random variable v, which is the average speed of the mobile users. Since the coverage area of the WLANs are relatively small and most of the users are almost static, v is taken only for pedestrians (e.g $E[v]$=2km/h). F is the length of the perimeter of cell (a WLAN with an arbitrary shape is assumed), and O is the area of the WLAN [9], [10], [11]. T_{ch} is the average packet holding time in the WLAN and follows an exponential distribution with mean $E[T_{ch}]$=$1/\mu_{ch}$. Since F is small for WLAN, it would be impractical to consider high speeds as the user will be out of range before they are able to use the connection. Hence in this study, near stationary speeds are considered. T can be defined as the total channel holding time of a packet, which is exponentially distributed with mean E[T] = $1/\mu$, where

$$\mu = \mu_{ch} + \mu_{wd} \tag{2}$$

Channel assignment schemes can be implemented in many different ways depending on the medium access control (MAC) and the system architecture. WLAN (the IEEE 802.11) MAC assures fair sharing of WLAN bandwidth/channels stations since WLANs use a shared medium [1, 4, 6, 7, 14, 19, 23]. Traffic load makes channels unreliable and unavailable for transmission. This causes severe degradation in the performance of WLANs. In order to solve the performance degradation problem in WLANs, frequency channels should be assigned to APs in an appropriate and efficient manner. Hence, in this paper, an analytical model developed maximizes the bandwidth usage according to the number of mobile users within the WLAN.

In other words, considering the behaviour of MAC mechanism, the users fairly share WLAN bandwidth. This sharing can be thought as a dynamic bandwidth allocation in WLAN. This is

done by (assuming the call is accepted with certain bandwidth assigned) optimal allocation of WLAN bandwidth for the number of channels required. The main objective of using dynamic bandwidth allocation is to improve the overall QoS of an entire WLAN by sharing the available bandwidth. When dynamic channel allocation scheme is employed, the channel does not change for different numbers of packets in the performance model employed.

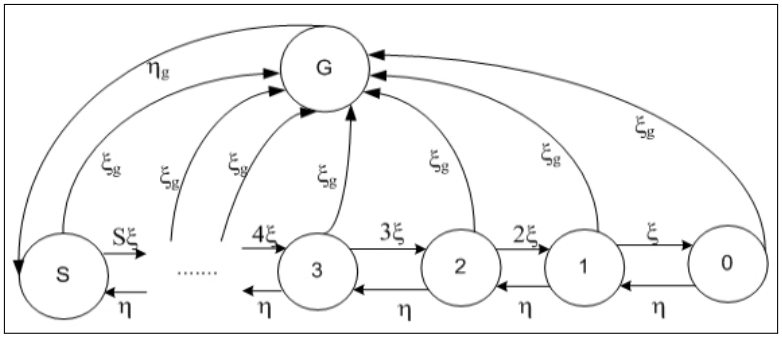


Figure 2: The operative states of the proposed system

The failure and repair behaviour of each channel are also represented by exponentially distributed failure and repair times with means $1/\xi$ and $1/\eta$, respectively [**?**], [16]. A single repairman facility is assumed for all of the channels but the models provided can easily be extended for systems with multiple repairmen facilities. An inoperative period of a channel would also include the possible waiting time for a repairman. No operative channel can be idle if there are packets waiting, and no repairman can be idle if there are failed channels waiting for repair. On the other hand, the system (AP) can fail, and the repair facility is provided with means $1/\xi_g$ and $1/\eta_g$, respectively. In case of global failures (ξ_g is the respective global failure rate) the priority of repair is given to the AP because the channels cannot be used when the AP is inoperative. All inter-arrival, service, failure/repair of channels and global failure and repair time variables are distributed exponentially and are independent of each other. The operative state of such a system is illustrated in Fig. 2. The states 1, 2, 3, $\cdots$, S are the working states of the multi-channel WLAN. In state 0 there are no channels available due to the channel failures and in state G there are no channels available due to the AP failures.

3 Two Dimensional Markov Representation of The Proposed Model

The state of the system at time t can be described by a pair of integer valued random variables, $I(t)$ and $J(t)$, specifying the channels plus global failure/recovery configurations and the number of packets present, respectively. Here, channel configuration, and hence the range of $I(t)$, refers to the operative states of the channels. In general, there are $S+2$ configurations, represented by the values $I(t) = 0, 1, \cdots, S, G$. The first $S+1$ configurations represent the number of available channels in the system (from 0 to S). The $(S+2)^{th}$ state is used to represent the system in case the AP is down (state G). When the AP is repaired with repair rate η_g, the system resets with S operative channels. $J(t)$ is the number of jobs in the system, $J(t) = 0, 1, \ldots, L$. $Z = [I(t), J(t)]$; $t \geq 0$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, $(0, 1, \cdots, S+1)$ x $(0, 1, \cdots, L)$.

Let the possible operative states of both channels and AP be represented in the horizontal direction (the number of channels available and G for the AP) and the number of packets in the vertical direction of a lattice strip. Here, A is the matrix of instantaneous transition rates from state (i, j) to state (k, j), (i=0, 1, $\cdots$, $S+1$; k=0, 1, $\cdots$, $S+1$; $i \neq k$; j=0, 1, $\cdots$, L), with zeros

on the main diagonal, caused by a change in the state. These are the purely lateral transitions of the process Z. Matrices B and C are transition matrices for one-step upward and one-step downward transitions respectively [9], [10]. B is the matrix of one-step upward transition rate from state (i, j) to state $(k, j+1)$, (i=0,1, $\cdots$, $S+1$; k=0,1, $\cdots$, $S+1$; and j=0,1, $\cdots$, L), caused by a packet arrival. Clearly, the elements of A depend on the parameters S, ξ, η, ξ_g, η_g. The transition matrices of a system with S channels are of size $(S+2)\times(S+2)$. It is possible to specify the numbering of the matrices as (0, 1, 2, 3, $\cdots$, $S+1$) for states (0, 1, 2, 3,$\cdots$, S, G) respectively. The state transition matrices A, A_j, B, and B_j, can be given as follows:

$$A = A_j = \begin{pmatrix} 0 & \eta & 0 & 0 & 0 & 0 & 0 & \xi_g \\ \xi & 0 & \eta & 0 & 0 & 0 & 0 & \xi_g \\ 0 & 2\xi & 0 & \eta & 0 & 0 & 0 & \xi_g \\ 0 & 0 & 3\xi & 0 & \ddots & 0 & 0 & \xi_g \\ 0 & 0 & 0 & \ddots & 0 & \eta & 0 & \xi_g \\ 0 & 0 & 0 & 0 & (S-1)\xi & 0 & \eta & \xi_g \\ 0 & 0 & 0 & 0 & 0 & S\xi & 0 & \xi_g \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta_g & 0 \end{pmatrix}$$

$$B = B_j = \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

On the other hand, C is the matrix of one-step downward transition rate from state (i, j) to state $(k, j-1)$, (i=0, 1, $\cdots$, $S+1$; k=0, 1, $\cdots$, $S+1$; and j=0, 1, $\cdots$, L), caused by the departure of a serviced packet. The transition rate matrices do not depend on j for $j \geq M$, where M is a threshold having an integer value. The elements of C and C_j depend on the parameter μ and can be given as follow:

$$C = C_j = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (S-1)\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & S\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Please note that the stations considered within the coverage are of the WLAN can be mobile. When such a scenario is considered, the transition rate matrix C depends on j since the departure rate caused by mobility depends on the number of packets in the queue [9, 10]. Therefore, the threshold M is taken as $M = L$. If the number of packets in the system is less than the number of available channels, a channel is assigned for each packet. Therefore, the downward transition rate is chosen as the minimum of number of packets and number of available channels. On the other hand, if the number of packets is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the departure rate μ_{wd} [9, 10] due to mobility. The matrix C is defined below, together with C_j matrices for two different regions explained above:

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu + W\mu_{wd} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & (S-1)\mu + W\mu_{wd} & 0 & 0 \\ 0 & 0 & 0 & 0 & S\mu + W\mu_{wd} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for j > S

$$C_j = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu + (j-S)\mu_{wd} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & (S-1)\mu + (j-S)\mu_{wd} & 0 & 0 \\ 0 & 0 & 0 & 0 & S\mu + (j-S)\mu_{wd} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for j ≤ S

$$C_j = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & min(1,j)\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & min(1,j)\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & min(1,j)\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & min(1,j)\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This system can be solved and the steady state probabilities, $p_{i,j}$ can be obtained using the steady state solution presented in the next section.

4 Steady State Solution

The solution is given for systems with bounded queueing capacities. The steady-state probabilities of the system considered can be expressed as:

$$p_{i,j} = lim_{t\to\infty}\ P(I(t) = i, J(t) = j); \tag{3}$$

$$0 \le i \le S+1,\ 0 \le j \le L$$

4.1 Spectral Expansion Solution

It is possible to use spectral expansion or matrix-geometric methods to solve the system considered. When matrix-geometric method is employed, a non-linear matrix equation is formed from the system parameters and an iterative method is employed in order to compute the minimal non-negative solution R of this equation. One of the main disadvantages of this method is that it is not possible know the number of iterations needed to compute R for a specified accuracy.

In this study spectral expansion method is employed for steady state solution. Let's define certain diagonal matrices of size $(S+2) \times (S+2)$ as follows:

$$D_j^A(i,i) = \sum_{k=0}^{S+1} A_j(i,k); \quad D^A(i,i) = \sum_{k=0}^{S+1} A(i,k); \tag{4}$$

$$D_j^B(i,i) = \sum_{k=0}^{S+1} B_j(i,k); \quad D^B(i,i) = \sum_{k=0}^{S+1} B(i,k); \tag{5}$$

$$D_j^C(i,i) = \sum_{k=0}^{S+1} C_j(i,k); \quad D^C(i,i) = \sum_{k=0}^{S+1} C(i,k); \tag{6}$$

and $Q_0 = B$, $Q_1 = A - D^A - D^B - D^C$, $Q_2 = C$.
all state probabilities in a row can be defined as:

$$v_j = (p_{0,j}, p_{1,j}, \cdots, p_{S+1,j}); j = 0, 1, 2 \ldots, L \tag{7}$$

The steady-state balance equations for bounded queuing systems ($0 \leq j \leq L$) can now be written as follows:

$$v_0[D_0^A + D_0^B] = v_0 A_0 + v_1 C_1 \tag{8}$$

$$v_j[D_j^A + D_j^B + D_j^C] = v_{j-1} B_{j-1} + v_j A_j + v_{j+1} C_{j+1}; \tag{9}$$

$$1 \leq j \leq M-1$$

$$v_j[D^A + D^B + D^C] = v_{j-1} B + v_j A + v_{j+1} C; \tag{10}$$

$$M \leq j \leq L$$

$$v_L[D^A + D^C] = v_{L-1} B + v_L A \tag{11}$$

The normalizing equation is given as follows:

$$\sum_{j=0}^{L} v_j e = \sum_{j=0}^{L} \sum_{i=0}^{S+1} P(i,j) = 1.0 \tag{12}$$

From the equations above, one can write

$$v_j Q_0 + v_{j+1} Q_1 + v_{j+2} Q_2 = 0 \tag{13}$$

$$(M-1) \leq j \leq (L-2)$$

Furthermore, the characteristic matrix polynomial $Q(\lambda)$ can be defined as:

$$Q_\lambda = Q_0 + Q_1 \lambda + Q_2 \lambda^2; \quad \bar{Q}_\beta = Q_2 + Q_1 \beta + Q_0 \beta^2; \tag{14}$$

where

$$\Psi Q_\lambda = 0; \quad |Q_\lambda| = 0; \quad \phi \bar{Q}_\beta = 0; |\bar{Q}_\beta| = 0; \tag{15}$$

λ and Ψ are eigenvalues and left-eigenvectors of Q_λ and β and ϕ are eigenvalues and left-eigenvectors of $\bar{Q}_\beta$, respectively. Note that, ϕ is a vector defined as

$$\phi = \phi_0, \phi_1, \ldots, \phi_{S+1} \tag{16}$$

and β as:

$$\beta = \beta_0, \beta_1, \ldots, \beta_{S+1}. \tag{17}$$

Furthermore, $v_j = \sum_{k=0}^{S+1}(a_k \Psi_k \lambda_k^{j-M+1} + b_k \phi_k(i) \beta_k^{L-j})$,M-1≤j≤ L and in the state probability form,

$$p_{i,j} = \sum_{k=0}^{S+1}(a_k \Psi_k \lambda_k^{j-M+1} + b_k \phi_k(i) \beta_k^{L-j}) \tag{18}$$

$$M - 1 \leq j \leq L$$

Where $\lambda_k (k = 0, 1, \ldots, S + 1)$ and $\beta_k (k = 0, 1, \ldots, S + 1)$ are $S + 2$ eigenvalues each, that are strictly inside the unit circle and a_k, b_k (k=0, 1,. . ., $S + 1$) are arbitrary constants which can be scalar or complex-conjugate [9], [10]. All the a_k, b_k values and the remaining v_j vectors can be obtained using the process in [9] and [10].

4.2 LU Factorization with Partial Pivoting

For the solution of real linear equations to obtain the scalars a_k, and b_k, (k=0, 1,. . ., $S + 1$) the set of linear equations are considered with a single right-hand side, using an LU factorization with partial pivoting, and iterative refinement [25]. LU factorization of the matrix employed is computed first with partial pivoting (L is lower triangular and U is unit upper triangular). An approximation for the answer vector x is then found by forward and backward substitutions. Using additional precision, the residual vector is calculated in turn. An iterative refinement approach is employed until full machine accuracy is obtained.

Once all the steady state probabilities $p_{i,j}$ (i=0, 1, . . ., $S + 1$; j=0, 1, . . ., L) are computed, a number of steady-state performance measures can be obtained. For numerical results and discussions, the mean queue length (MQL) and blocking probabilities (P_B) of multi-channel WLAN are considered as:

$$MQL = \sum_{j=0}^{L} j \sum_{i=0}^{S+1} p_{i,j} \tag{19}$$

$$P_B = \sum_{i=0}^{S+1} p_{i,L} \tag{20}$$

5 Numerical Results and Discussions

Numerical results are presented in this section for the performability analysis of multi-channel WLANs with global failures. The exact spectral expansion solution is employed to obtain performability measures. Numerical results are presented for performability measures of multi-channel WLAN (e.g. MQL and P_B). The parameters used are mainly taken from the relevant literature and are $E[T_{ch}] = 60sec$, E[v]=2km/h, R=100m [7, 8, 10, 11, 13, 18].

The results in Fig. 3 show that for low channel failure rates, the system can perform as good as systems without failures. However, as the channel failure rate increases the performance degradation becomes more evident (for ξ=0.1, the MQL values become up to twice as much as the MQL values calculated for the other rates). The MQL and P_B measures are presented in Fig. 4 and Fig. 5, respectively. The parameters used are the same as previous computations with various λ values.

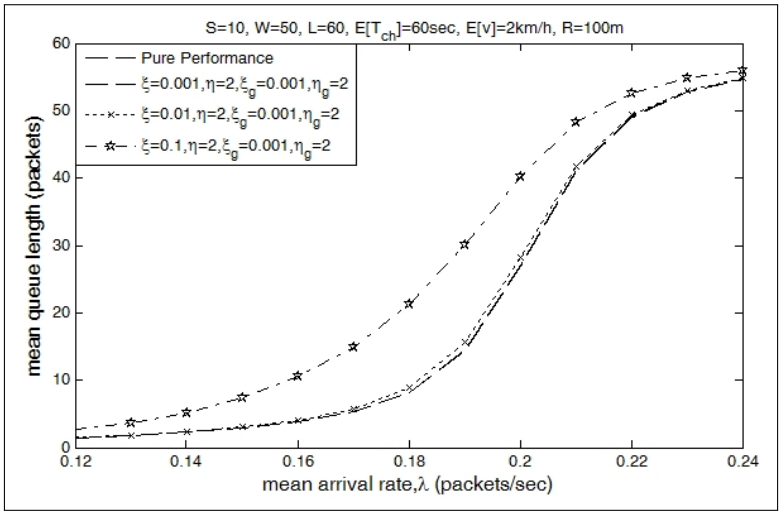


Figure 3: MQL results for systems with various channel failure rates

Fig. 4 and Fig. 5 show that the global failures (AP failures) affect the system significantly. The effects of global failures are more evident for blocking probabilities, since as mean arrival rate and global failure rate increase, the limiting effect of queue capacity becomes more evident. The effects of global failures are considered in Fig. 6 as well. This time a 25 channels system is considered (S=25) together with W=100. MQL values are presented as a function of mean arrivals. The other parameters used are same as previous computations.

Fig. 6 shows that when failure rate of the access point is taken as 0.1 the difference in MQL values is quite significant. However for systems with ξ_g=0.01/h and ξ_g =0.001/h the difference is not as significant. This is mainly due to good repair facilities provided. Since the η_g=2/h the mean repair time is taken as half an hour for the access point. The effects of having various repair times are considered in Fig. 7. The other parameters considered are the same as the ones used for Fig. 6.

The repair time of the access point affects the performance of the system significantly. Fig. 7 clearly shows that systems with mean repair time equal to one hour performs significantly better then systems with $1/\eta_g$=1.5h and $1/\eta_g$=2h. In the model considered the priority of repair is given to the access point. However, since the repair time also includes delay factors such as the time needed for transportation of the repairman, configuration of access point etc., the repair facility should be carefully considered in order to meet expected levels of performability. The results demonstrate that the systems may become overwhelmed because of long repair times,

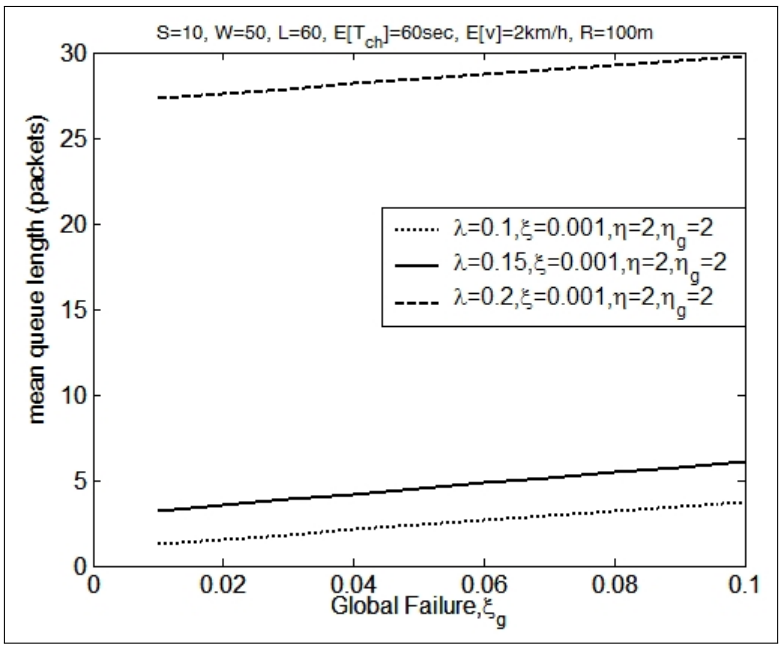


Figure 4: MQL results for systems with various global failure rates

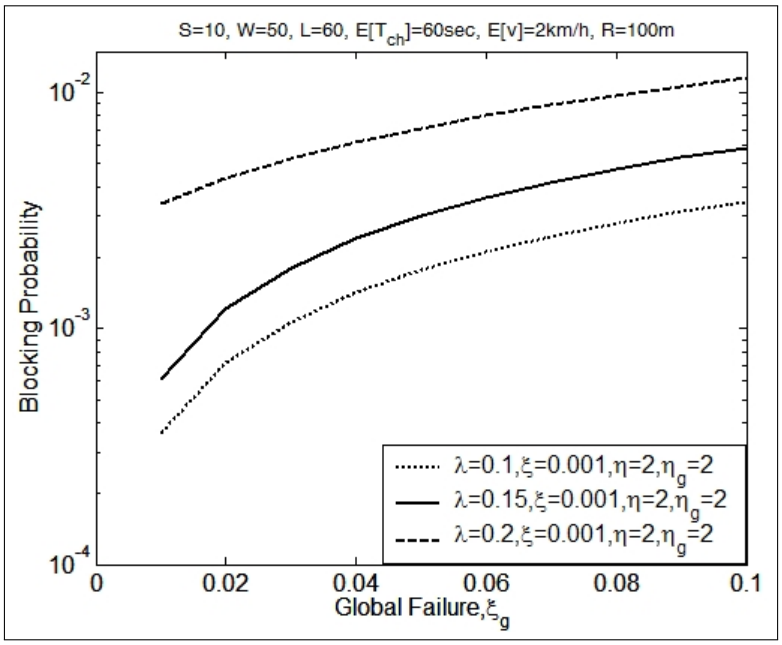


Figure 5: Blocking probabilities for systems with various global failure rates

especially if heavy loads of packet requests are expected.

Fig. 8 shows the effects of having various numbers of channels in a WLAN with various mean incoming arrival rates. The other parameters used are the same as the ones used for Fig. 7. Fig. 8 shows that the number of channels affects the overall performability of the system significantly. When λ=0.25 is considered, the MQL takes values close to L, such as 104.32 for five channel systems and it is around 1.042 for systems with 25 channels. Systems with different number of available channels may perform similarly for heavy loads and light loads, however, it is obvious that as the number of channels employed increases, the system performs better.

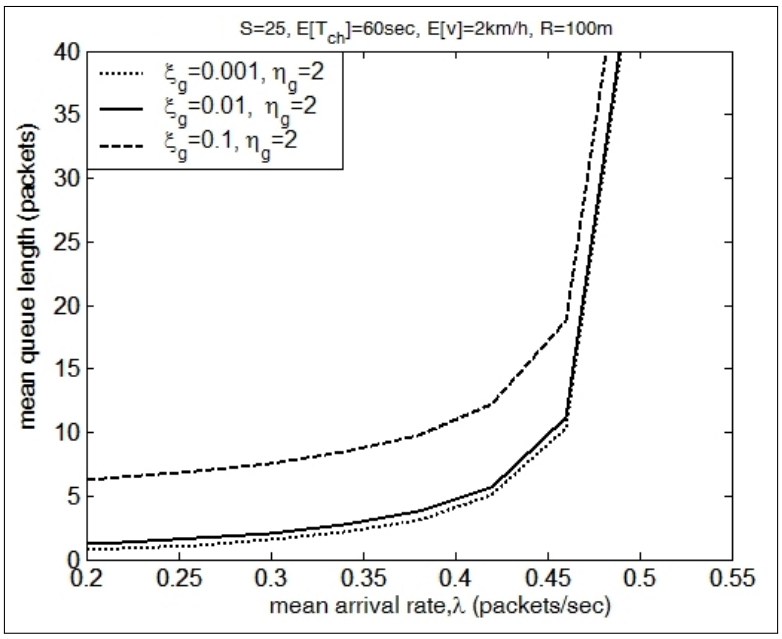


Figure 6: MQL as a function of λ for S=25 and W=100 for various ξ_g values

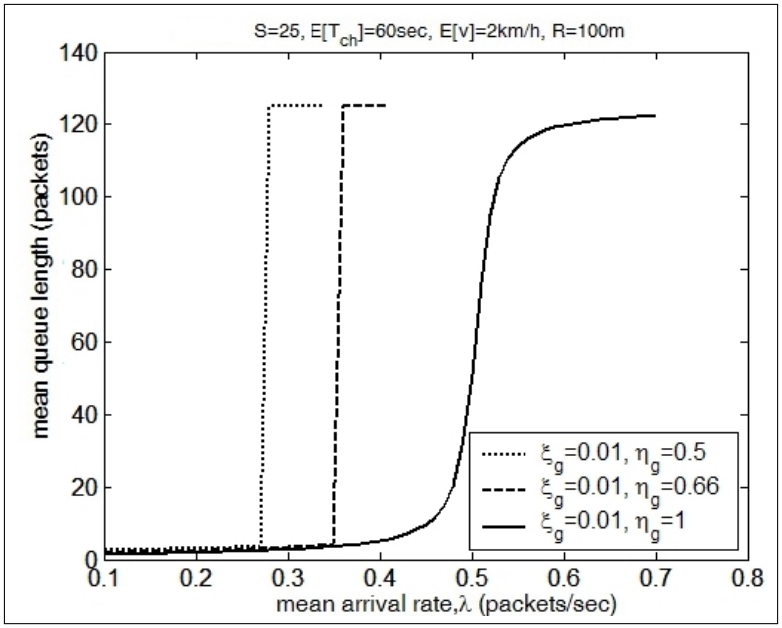


Figure 7: MQL as a function of λ for S=25 and W=100 for various η_g values

Finally, Fig. 9 shows the MQL as a function of mean rate for incoming packet requests. Two different scenarios are considered for the computations. In the first scenario the mobile stations waiting in the queue for channels are static, whereas in the second scenario the mobile stations are pedestrians with E[v]=2km/h. The other parameters are S=10, W=50, ξ=0.01/h, η=2/h, ξ_g=0.001/h, η_g=2/h. The results show that, although the velocity of pedestrian users is relatively low, in case of high arrival rates, the mean number of users leaving the system due to mobility is quite significant. For example, for λ=0.2 MQL is 28.3 for systems with static users in the queue, and 9.92 for systems with pedestrian users in the queue.

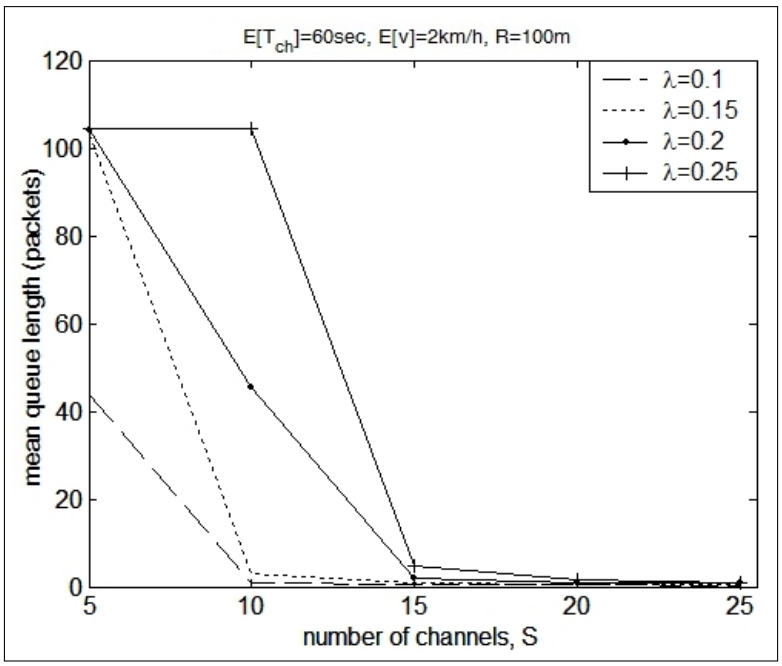


Figure 8: MQL as a function of S for various λ values

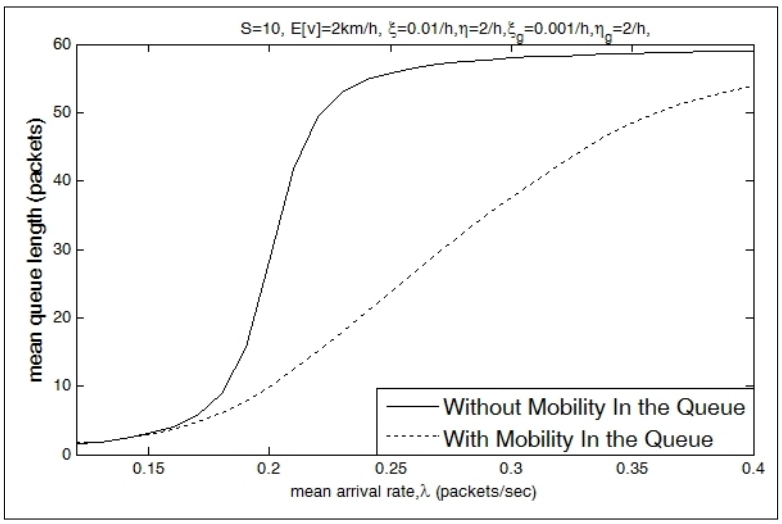


Figure 9: MQL as a function of λ, for mobile users waiting for service

6 Conclusions

In this paper, multi-channel WLAN is considered for performability evaluation. A two dimensional modelling approach and an exact solution approach where iterative refinements are used to increase the accuracy for solution of system of linear equations are employed. The multi-channel failures/repairs, and global failures/repairs of WLANs modelled for exact solution. The dynamic channel allocation scheme is also considered in an attempt to maximise the performance of the underlying infrastructure. The proposed model is used to analyse QoS measures such as mean queue length and blocking probabilities. The presented examples were kept simple due to the introductory nature of the proposed model for the multi-channel WLANs. Obviously, the analysis can be expanded upon for more informed decisions, since the multi-channel WLANs are common today. Recent WLANs have multiple wireless radios each operating independently on different channels. The analysis of the multi-channel WLANs is an important issue in order to achieve better QoS measurements in future wireless and communication systems.

In this paper, in order to obtain more realistic QoS measures, considerable amount of focus is given operative states of multi-channel as well as global failures of WLANs with the proposed analytical model and its exact solution. The results mainly show that, channel and global failures affect the QoS of the system significantly. Therefore, the system availability is important for the future system design and modelling. Results also show that the failures cause significant performance degradations in the system, and the channel and global recovery are important parameters for the system's performance. The systems with higher number of channels perform better as expected. As is widely agreed, future mobile communication systems, fourth-generation (4G) systems, will be heterogeneous networks, which provide ubiquitous access and seamless mobility across heterogeneous network technologies, such as cellular, WLAN, and broadcast access. In 4G wireless networks, mobility management is still one of the most important issues that need to be considered. Thus, results provided in Fig. 9 are very important in that sense.

The results mainly show the analytical model and the exact steady state solution approach presented is very useful in future systems since the recent WLANs technologies operate in multi-channel basis (e.g Hiperlan2 or WiMAX). The model presented in this paper is flexible and can be adopted to the other WLANs.

Bibliography

[1] M. Gong and C. Williamson (2009); Scheduling Issues in Multi-Channel Wireless Networks, *IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'09)*, 1-10.

[2] Z. Hua, M. Li, I. Chlamtac, and B. Prabhakaran (2004); A survey of quality of service in IEEE 802.11 networks, *IEEE Wireless Communications Journals*, 11: 6-14. 2004.

[3] H. Bobarshad (2010); Evaluation and Modelling of Real Time Video Transmission Over WLAN, PhD thesis, Department of Electronic Engineering, King's College, London.

[4] Q. Wang, M. Liu (2010); Throughput Optimal Switching in Multi-channel WLANs, *IEEE Transactions on Mobile Computing*, 99: 383-388.

[5] P. Bahl, A. Adya, J. Padhye, and A. Wolman (2004); Reconsidering Wireless Systems with Multiple Radios, *ACM Computer Communication Review*, 34(5):39-46.

[6] G. Bianchi (2000); Performance analysis of the IEEE 802.11 distributed coordination function, *IEEE Journal on Selected Areas on Communications*, 18(3): 535-547.

[7] R. P. Liu, G. J. Sutton, and I. B. Collings (2010); A New Queuing Model for QoS Analysis of IEEE 802.11 DCF with Finite Buffer and Load, *IEEE Transaction on Wireless Communications*, 9(8): 2664- 2675.

[8] J. Yin, W. Xiaodong, and D. P. Agrawal (2005); Modeling and optimization of wireless local area network, *IEEE Computer Communications*, 28(10): 1204-1213.

[9] E. Ever, Y. Kirsal, and O. Gemikonakli (2009); Performability Modelling of Handoff in Wireless Cellular Networks and the Exact Solution of System Models with Service Rates Dependent on Numbers of Originating and Handoff Calls, *International Conference on Modelling and Simulation*, Czech Republic, 282-287.

[10] Y. Kirsal, E. Ever, O. Gemikonakli, G. Mapp (2011); Modelling and Performability Analysis of WLANs as a Queuing Model with Channel/Access Point Failures and Reconfiguration, *IEEE Proceedings of 5th European Modelling Symposium (EMS 2011)*, Universidad Politecnica de Madrid, 440-445.

[11] Y. Ma, J. Han, and K.S. Trived (2001);, Composite Performance and Availability Analysis of Wireless Communication Networks, *IEEE Trans. on Vehicular Technology*, 50(5): 1216-1223.

[12] F. Capar, T. Weiss, I. Martoyo, and F. Jondral (2003); Analysis of Coexistence Strategies for Cellular and Wireless Local Area Networks, *Vehicular Technology Conference*, DOI:10.1109/VETECF.2003.1285338, 3: 1812 - 1816.

[13] Gowrishankar, G.N. Sekhar, and P.S. Satyanarayana (2009); Analytic Performability Model of Vertical Handoff in Wireless Networks, *Journal of Computer Science*,5(6):445-450.

[14] J. So, and N. Vaidya (2004); Multi-channel MAC for ad hoc networks: handling multi-channel hidden terminals using a single transceiver, *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, 222-233.

[15] B. Paramvir, A. Adya, J. Padhye, and, A. Walman (2004); Reconsidering wireless systems with multiple radios, *SIGCOMM Computer Communications Review*, 34(5): 9-46.

[16] D. Chen, C. Kintala, S. Garg, and K. S. Trivedi (2003); Dependability enhancement for IEEE 802.11 wireless LAN with redundancy techniques, *International Conference on Dependable Systems and Networks*, 521-528.

[17] A.V. Nikolov (2008); Analytical Model For a Multiprocessor With Private Caches And Shared Memory, *International Journal of Computers Communications & Control*, 3(2): 172-182, 2008.

[18] W. Xia, and, L. Shen (2007); Modeling and Analysis of Handoffs in Cellular and WLAN integration, *IEEE International Conference on Communications*, 385-391.

[19] Y. Zhao, and K. K. Leung (2006); Adaptive channel allocation for IEEE 802.11 wireless LANs, *The* 12^{th} *European Wireless 2006 (EW2006) Conference*, Athens, 1-6.

[20] A. Amer and F. Gebali (2009); General model for single and multiple channels WLANs with quality of service support, *International Journal of Wireless and Mobile Networks*, 1(2): 1-19.

[21] T. Kuang, Q. Wu and C. Williamson (2005); MRMC: A Multi-Rate Multi-Channel MAC Protocol for Multi-Radio Wireless LANs, *Proc. of the 2005 Workshop on Wireless Networks and Communication Systems (WiNCS)*, Philadelphia, 263-272.

[22] A. Raniwala and T. Chiueh (2005); Architecture and Algorithms for an IEEE 802.11-Based Multi-Channel Wireless Mesh Network, *Infocom*, DOI:10.1109/INFCOM.2005.1498497, 3: 2223- 2234.

[23] V. Gupta, M. Gong, S. Dharmaraja, C. Williamson (2011); Analytical modeling of bidirectional multi-channel IEEE 802.11 MAC protocol, *Int. J. Communication Systems*, 24(5): 647-665.

[24] S.H. Kim, D.W. Kim, Y.J. Suh (2013); A survey and comparison of multichannel protocols for performance anomaly mitigation in IEEE 802.11 wireless networks. *Int. J. Communication Systems*, 26(10): 1288-1307.

[25] Numerical Algorithms Group (2005), *NAG C Library Manual*, NAGNP3660/8 December 2005 ISBN:978-1-85206-206-4 (ISBN 1-85206-206-1).

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):567-578, August, 2015.

Robust Adaptive Self-Organizing Wavelet Fuzzy CMAC Tracking Control for Deicing Robot Manipulator

T.Q. Ngo, T.V. Phuong

ThanhQuyen Ngo*
Faculty of Electrical Engineering,
Industrial University of HCM City , HCM City, Vietnam
*Corresponding author: thanhquyenngo2000@yahoo.com

TaVan Phuong
Faculty of Electrical and Electronics Engineering,
HCMC University of Technology And Education, Vietnam
tavphuong@yahoo.com; phuongtv@hcmute.edu.vn

Abstract:
In this paper, a robust adaptive self-organizing control system based on a novel wavelet fuzzy cerebellar model articulation controller (WFCMAC) is developed for an n-link robot manipulator to achieve the high-precision position tracking. This proposed controller consists of two parts: one is the WFCMAC approach which is implemented to cope with nonlinearities, due to the novel WFCMAC not only incorporates the wavelet decomposition property with fuzzy CMAC fast learning ability but also it will be self-organized; that is, the layers of WFCMAC will grow or prune systematically. Therefore, dimension of WFCMAC can be simplified. The second is the order which is the adaptive robust controller which is designed to achieve robust tracking performance of the system. The adaptive tuning laws of WFCMAC parameters and error estimation of adaptive robust controller are derived through the Lyapunov function so that the stability of the system can be guaranteed. Finally, the simulation and experimental results of novel three-link deicing robot manipulator are applied to verify the effectiveness of the proposed control methodology.
Keywords: Wavelet, CMAC, Deicing robot manipulator.

1 Introduction

In general, robotic manipulators have to face various uncertainties in their dynamics, such as friction and external disturbance. It is difficult to establish an exactly mathematical model for the design of a model-based control system. In order to deal with this problem, the fuzzy logic control (FLC) has found extensive applications for complex and ill-defined plants [1-2], and is suitable for simple second order plants. However, in case of complex higher order plants, all process states are required as fuzzy input variables to implement state feedback FLCs. All the state variables must be used to represent contents of the rule antecedent. Therefore, it requires a huge number of control rules and much effort to create them. The neural networks (NNs) are a model-free approach, which can approximate a nonlinear function to arbitrary accuracy [3–4]. However, the learning speed of the NNs is slow, since all the weights are updated during each learning cycle. In addition, the fully connected NNs are sensitive to training data. Therefore, the ability of function approximation of a general multiplayer NN is restricted to requiring online learning.

Based on the advantages of fuzzy and neural networks, the fuzzy neural network (FNN) which incorporates advantages of fuzzy inference and neuron-learning has been developed and its effectiveness is demonstrated in solving control problems [5–6]. Recently, many applications have been implemented quite successfully based on wavelet neural networks (WNNs) which combine

the learning ability of network and capability of wavelet decomposition property [7–8]. Different from conventional NNs, the membership functions of WNN are spatially localized wavelet functions; and therefore, the WNNs are capable of learning more efficiently than conventional NNs for control and system identification as has been demonstrated in [7]. As a result, WNNs have been used to deal with uncertainties and nonlinearity in [8].

To overcome the disadvantages of NNs, cerebellar model articulation controller (CMAC) was proposed by Albus in 1975 [9] for the identification and control of complex dynamical systems, due to its advantage of fast learning property, good generalization capability and ease of implementation by hardware [10–11]. Conventional CMACs are regarded as nonfully connected perceptron-like associative memory network with overlapping receptive fields which used constant binary or triangular functions. The disadvantage is that their derivative information is not preserved. To acquire the derivative information of input and output variables, Chiang and Lin [12] developed a CMAC network with a differentiable Gaussian receptive-field basis function and provided the convergence analysis for this network. The advantages of using CMAC over neural network in many applications were well documented [13–14]. However, in the above CMAC studies, the structure of CMAC are not merited of the high-level human knowledge representation and thinking of fuzzy theory.

In this paper, we propose a novel robust adaptive self-organizing wavelet fuzzy CMAC (RASOWFCM) control system for three-link deicing robot manipulator to achieve the high-precision position tracking. This control system combines the advantages of fuzzy inference system with CMAC and wavelet decomposition capability and it does not require prior knowledge of a certain amount of memory space, and a self-organizing approach of this control system demonstrates the properties of generating and pruning the input layers automatically The developed self-organizing rule of WFCMAC is clearly and easily used for real-time systems and the adaptive robust controller which are designed to achieve robust tracking performance of the system. This is the first contribution of this paper. The second contribution is that it proposes novel architecture and mathematical model of deicing robot which can be effective in practical applications. The adaptive tuning laws of WFCMAC parameters and error estimation are derived in Lyapunov method.

This paper is organized as follows: system description is given in Section 2. The structure of self-organizing WFCMAC and RASOWFCM control system are presented in Sections 3 and 4. Numerical simulation and experimental results of a three-link deicing robot manipulator under the possible occurrence of uncertainties are provided to demonstrate the tracking control performance of the proposed RASOWFCM system in Section 5. Finally, conclusions are drawn in Section 6.

2 System description

2.1 Robotic dynamic model

In general, the dynamics of an n-link robot manipulator may be expressed in the following Lagrange form:

$$M(q)\ddot{q} + V(q,\dot{q})\dot{q} + G(q) + F(\dot{q}) + \tau_{\mathrm{d}} = \tau, \tag{1}$$

where $q, \dot{q}, \ddot{q} \in \mathbb{R}^n$ are the joint position, velocity and acceleration vectors, respectively, $M(q) \in \mathbb{R}^{n\times n}$ denotes the inertia matrix, $V(q,\dot{q}) \in \mathbb{R}^{n\times n}$ expresses the matrix of centripetal and Coriolis forces, $G(q) \in \mathbb{R}^{n\times 1}$ is the gravity vector, $\tau_{\mathrm{d}} \in \mathbb{R}^{n\times 1}$, $F(\dot{q})$ represents the vector of external disturbance, friction term, respectively, and $\tau \in \mathbb{R}^{m\times 1}$ is the torque vectors exerting on joints.

In this paper, a novel three-link deicing robot manipulator, as shown in Fig.1, is utilized to verify dynamic properties which are given in Section 5.

In our controller design, the following properties will be used [15].

Property 1: The inertia matrix $M(q)$ is symmetric and positive definite. It is also bounded as a function of $q : m_1 I \leqslant M(q) \leqslant m_2 I, m_1, m_2 > 0$.

Property 2: $\dot{M}(q) - 2V(q, \dot{q})$ is a skew symmetric matrix. Therefore, $y^{\mathrm{T}}[\dot{M}(q) - 2V(q, \dot{q})]y = 0$, where y is a $n \times 1$ nonzero vector. The control problem is to force $q(t) \in \mathbb{R}^n$ to track a given bounded reference input signal $q_{\mathrm{d}}(t) \in \mathbb{R}^n$. we have the tracking error as follows:

$$e = q_{\mathrm{d}}(t) - q(t), \tag{2}$$

and the filtered system tracking error vector is defined as

$$r = \dot{e} + \Lambda e, \tag{3}$$

where $\Lambda = \Lambda^{\mathrm{T}} > 0$, by the differentiating $r(t)$ with respect to e and using (1), the arm dynamics can be written in terms of the filtered tracking error vector as follows:

$$\begin{aligned} M(q)\dot{r} &= -V(q, \dot{q})r - \tau + M(q)(\ddot{q}_{\mathrm{d}} + \Lambda\dot{e}) + V(q, \dot{q})(\dot{q}_{\mathrm{d}} + \Lambda e) + F(\dot{q}) + G(q) + \tau_{\mathrm{d}} \\ &= -V(q, \dot{q})r - \tau + f(x), \end{aligned} \tag{4}$$

where the nonlinear function $f(x)$ is defined as follows:

$$f(x) = M(q)(\ddot{q}_{\mathrm{d}} + \Lambda\dot{e}) + V(q, \dot{q})(\dot{q}_{\mathrm{d}} + \Lambda e) + G(q) + F(\dot{q}) + \tau_{\mathrm{d}}, \tag{5}$$

where $x \triangleq [e^{\mathrm{T}} \quad \dot{e}^{\mathrm{T}} \quad q_{\mathrm{d}}^{\mathrm{T}} \quad \dot{q}_{\mathrm{d}}^{\mathrm{T}} \quad \ddot{q}_{\mathrm{d}}^{\mathrm{T}}]$ denotes the variables of the nonlinear function.

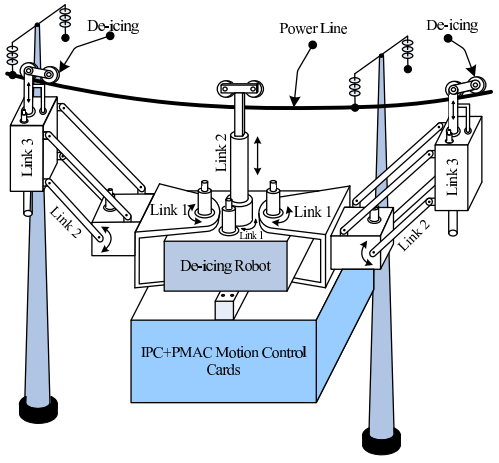


Figure 1: Architecture of three-link deicing robot manipulator.

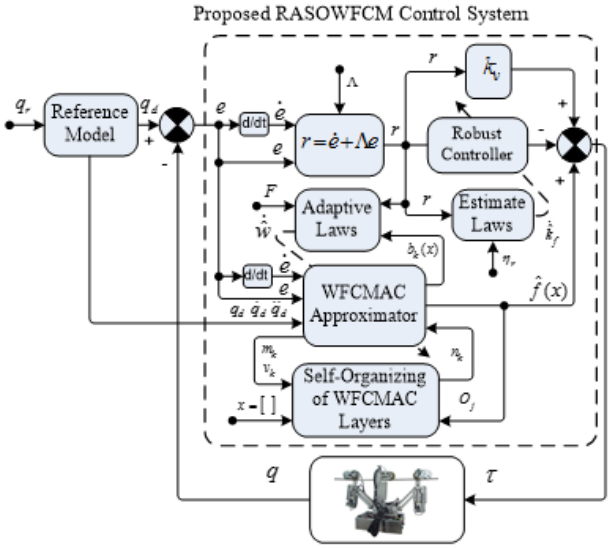


Figure 2: Block diagram of RASOWFCM control system.

2.2 Defined control law

Now, we define a control input torque as follows:

$$\tau_0 = \hat{f}(x) + K_{\mathrm{v}} r \tag{6}$$

with $\hat{f}(x)$ an estimate of $f(x)$ and a gain matrix $K_{\mathrm{v}} = {K_{\mathrm{v}}}^{\mathrm{T}} > 0$. From (4), the closed-loop system becomes

$$M(q)\,\dot{r} = -(K_{\mathrm{v}} + V(q,\dot{q}))\,r + \tilde{f}(x), \tag{7}$$

where functional estimated error is given by

$$\tilde{f}(x) = f(x) - \hat{f}(x). \tag{8}$$

This is a system error wherein filtered tracking error is driven by this estimated error function. The control τ_0 incorporates a proportional plus derivative (PD) term in $K_{\mathrm{v}}r = K_{\mathrm{v}}(\dot{e}+\Lambda e)$. In the remainder of the paper we shall use (7) to focus on selecting WFCMAC weight tuning algorithms that guarantee the stability of the filtered tracking error $r(t)$. Then, since (3), with the input considered as $r(t)$ and the output as $e(t)$, describes a stable system, standard techniques [16] guarantee that $e(t)$ exhibits stable behaviour. In fact, $\|e\|_2 \leqslant \|r\|_2/\sigma_{\min}(\Lambda)$, $\|\dot{e}\|_2 \leqslant \|r\|_2$ with $\sigma_{\min}(\Lambda)$ the minimum singular value of Λ. Generally, Λ is diagonal so that $\sigma_{\min}(\Lambda)$ is the smallest element of Λ.

3 Structure of WFCMAC

3.1 Brief of the WFCMAC

The main difference between the FCMAC and the original CMAC is that the association layer in the FCMAC is the rule layer which is represented as follows:

$$\begin{aligned} \mathbb{R}^{\lambda}: \quad & \text{if } x_1 \text{ is } \mu_{1k}, \text{ and } x_2 \text{ is } \mu_{2k}, \ldots, x_i \text{ is } \mu_{ik}, \text{Then } o_{ij} = w_{ij} \\ & \text{for } k = 1,2,\ldots,n_k, i = 1,2,\ldots,n, \lambda = 1,2,\ldots,n_\lambda \text{ and } j = 1,2,\ldots,m_j, \end{aligned} \tag{9}$$

where n is the input dimension, n_k is the number of layers for each input dimension, the number of rules n_λ equals the number of layers and m_j is the output dimension, μ_{ik} is the fuzzy set for the ith input, the kth layer, and w_{jk} is the output weight in the consequent part. Based on [17], a novel WFCMAC is represented and shown in Fig.4. It is combined a wavelet function with the FCMAC consists of input, association memory, receptive field, and output spaces, is proposed to implement the WFCMAC estimate in RASOWFCM control system which is shown in Fig.2. The signal propagation is introduced according to functional mapping as follows:

1. The first mapping X: $X \rightarrow A$ we assume that each input state variable $x = [\ x_1 \quad x_2 \quad \cdots \quad x_n\ ]$ can be quantized into n_e discrete states and that the information of a quantized state is regarded as a wavelet receptive-field basic function for each layer. The mother wavelet is a family of wavelets. The first derivative of basic Gaussian function for each layer is given here as the mother wavelet which can be represented as follows:

$$\mu_{ik}(x_i) = -\frac{x_i - m_{ik}}{\sigma_{ik}} \exp\left[-\frac{(x_i - m_{ik})^2}{2\sigma_{ik}^2}\right], \qquad i = 1,2,\cdots,n, k = 1,2,\cdots,n_k, \tag{10}$$

where μ_{ik} represents the kth layer of the input with m_{ik} is a translation parameter and σ_{ik} is dilation.

2. The second mapping A: $A \rightarrow R$ the information μ_{ik} of each kth layer relates to each location of receptive field space. Fig.4 illustrates a structure of two-dimension WFCMAC with wavelet basic function and $n_e = 7$ case. Areas of receptive field space are formed by multiple-input regions are called hypercube; i.e in the fuzzy rules in (9), the product is used as the "and" computation in the consequent part. The firing of each state in each layer can obtain the weight

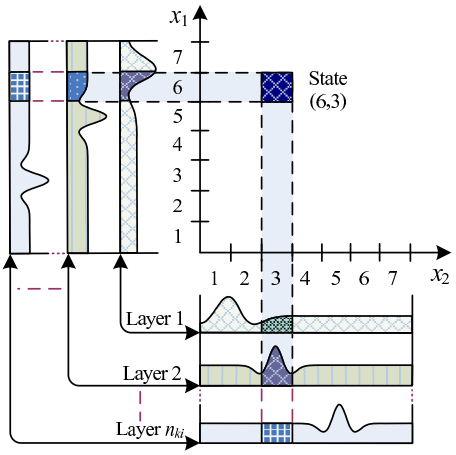


Figure 3: Two-dimension WFCMAC with wavelet function and $n_e = 7$.

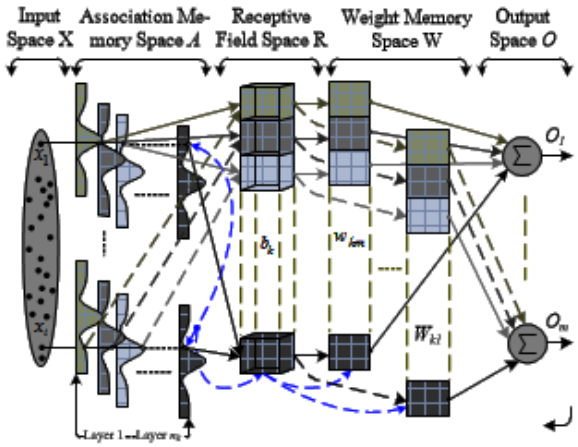


Figure 4: Block diagram of WFCMAC.

of each corresponding layer. Assuming that in 2-D WFCMAC case shown in Fig.3, input state vector is (6,3), the content of kth hypercube can be obtained as follows:

$$b_k(x,\, m_k,\, \sigma_k) = \prod_{i=1}^{n} \mu_{ik}(x_i). \tag{11}$$

3. Output mapping O: The WFCMAC output is an algebraic sum of activated weighs with hybercube elements. The jth output mathematic form can be expressed as follows:

$$O_j = \begin{bmatrix} w_{j1} & w_{j2} & \cdots & w_{jn_k} \end{bmatrix} \begin{bmatrix} b_1(x) \\ b_2(x) \\ \vdots \\ b_{n_k}(x) \end{bmatrix} = \sum_{k=1}^{n_k} w_{jk} \prod_{i=1}^{n} \mu_{ik}(x_i) \quad j = 1, 2, \cdots,\, m_j. \tag{12}$$

The block diagram shows in Fig.2, in which the WFCMAC plays a major role in the nonlinear function estimation. Because, there is a trade-off between a designed performance and a tedious computation so we must choose a reasonable number of layers. However, if the number of layers chosen is too small, the learning process may be insufficient to achieve a desired performance. Otherwise, if the number of layers chosen is too large, the calculation burden becomes too heavy, so it is not suitable for real-time applications. To deal with this problem, a self-organizing WFCMAC is proposed includes structure and parameter learning.

3.2 Self-organizing WFCMAC

In this section, It is necessary to determine a structure learning in order to add a new layer whether to add a new layer in membership layers depends on the firing strength $b_k \in \mathbb{R}^{n_k}$ of each layer for each incoming data x_i. If the firing strength $b_k \in \mathbb{R}^{n_k}$ of each layer for new input data x_i falls outside the bounds of the threshold, then, the WFCMAC approach will generate a new layer. The self-organizing WFCMAC can be also summarized based on [17] as follows:

1. Calculate a distance of mean MD_k, $k = 1, 2, \cdots, n_k$ in association memory A as follows:

$$MD_k(x) = \|x - |m_k|\|_2 \quad Where \quad m_k = \begin{bmatrix} m_{1k} & m_{2k} & \cdots & m_{nk} \end{bmatrix} \tag{13}$$

2. Using Max-Min method is proposed for layer growing. Find

$$\hat{k} = \arg \min_{1 \leqslant k \leqslant n_k} MD_k(x), k = 1, 2, \cdots n_k m, \quad if \quad \max_i MD_{\hat{k}}(x) > K_g, \tag{14}$$

Here K_g is a threshold value of adaptation with $0 < K_g \leqslant 1$. In our case $K_g = 0.1$ and a new layer is generated. This means that for a new input data, the exciting value of existing basic function is too small. In this case, number of layers increased as follows:

$$n_k(t+1) = n_k(t) + 1, \tag{15}$$

Where n_k is the number of layers at time t. Thus, a new layer will be generated and then the corresponding parameters in the new layer such as the initial mean and variance of Gaussian basic function in association memory space and the weight memory space will be defined as

$$m_{in_k} = x_i, \quad \sigma_{in_k} = \sigma_{\hat{k}}, \quad w_{in_k} = 0. \tag{16}$$

Another self-organizing learning process is considered to determine whether to delete existing layer, which is inappropriate. A Max-Min method is proposed for layer pruning.

Considering the jth output of WFCMAC in (12), the ratio of the kth component to jth output is defined as

$$MM_{jk} = \frac{v_{jk}}{O_j}, \quad k = 1, 2, \cdots, n_k. \tag{17}$$

Where $v_{jk} = w_{jk} b_k(x)$, Then, the minimum ratio of the kth component is defined as follows:

$$\tilde{k} = \arg \min_{1 \leqslant k \leqslant n_k} \max_{1 \leqslant j \leqslant m} MM_{jk}, \qquad if \quad MM_{\tilde{k}} \leqslant K_c. \tag{18}$$

Here K_c is a predefined deleting threshold. In our case $K_c = 0.03$ and the $\tilde{k}$th layer will be deleted. This means that for an output data, if the minimum contribution of a layer is less than the deleting threshold, then this layer will be deleted.

4 RASOWFCM Design

In the RASOWFCM scheme is shown in Fig.2, the WFCMAC is used to estimate an unmodelled nonlinear function, Moreover, the RASOWFCM law and adaptive tuning algorithms for WFCMAC are introduced from the stability analyses of the closed-loop system by using Lyapunov method Input of the WFCMAC estimator are the elements in the filtered error vector and joint positions signal, Output of the WFCMAC estimator are nonlinear dynamic function vectors in the local models.

Based on the powerful approximation ability [18], there exists an optimal WFCMAC estimator to approximate the nonlinear dynamic function in (5) such that

$$f(x) = W^{\mathrm{T}} b(x) + \varepsilon(x) \tag{19}$$

With W the ideal weight matrix and the estimated error vector $\varepsilon(x) \in \mathbb{R}^{n \times 1}$ are assumed to be given by

$$\tilde{W} = \arg \min_{\hat{W} \in M_w} [\sup_{x \in M_x} \| f(x) - \hat{W}^{\mathrm{T}} b(x) \|], 0 \leqslant \| \varepsilon \|_1 \leqslant k_f. \tag{20}$$

In which $\|\bullet\|$ is the Euclidean norm, M_x, M_w and k_f are the predefined compact sets of x and $\hat{W}$, and the positive constant

Define the WFCMAC functional estimate by

$$\hat{f}(x) = \hat{W}^{\mathrm{T}} b(x), \tag{21}$$

With $\hat{W}$ being the current values of the WFCMAC weight provided by the tuning algorithm. With the ideal weights required in (19) define the weight deviations or weight estimation errors as

$$\tilde{W} = W - \hat{W}. \tag{22}$$

With τ_o defined by (6), we design the control law as follow:

$$\tau = \tau_0 - \tau_r. \tag{23}$$

Where τ_0 is the main controller in (6), which consists of two terms: one is the PD controller and the other is WFCMAC controller is used to estimate the $f(x)$ nonlinear function and the adaptive robust controller τ_r is utilized to compensate for the approximation error between uncertainly model and WFCMAC approach.

We assume that the error bound is a constant during the observation. However, in practical application it is difficult to determine. Therefore, a bound estimation is developed to estimate this error bound. The estimation of error bound is defined as follows:

$$\tilde{k}_f = k_f - \hat{k}_f. \tag{24}$$

Where $\hat{k}_f$ is the estimated value of k_f. The adaptive robust controller is designed to compensate for the effect of the approximation error is selected as follows:

$$\tau_r = -\hat{k}_f \mathrm{sgn}(r). \tag{25}$$

Substituting (23) into (4) and using (19), (21) and (25), then, the closed-loop filtered error dynamics becomes:

$$M\dot{r} = -\left(K_{\mathrm{v}} + V(q,\, \dot{q})\right) r + \tilde{W}^{\mathrm{T}} b(x) + \varepsilon - \hat{k}_f \mathrm{sgn}(r). \tag{26}$$

Theorem 1. *Consider an n-link robot manipulator represented (1). If the control law of RASOWFCM is designed as (23), and the weight update law of WFCMAC and the adaptive law of the error bound are designed as (7), then the stability of the proposed RASOWFCM system can be ensured*

$$\dot{\hat{W}} = F\, b(x)\, r^{\mathrm{T}}, \quad \dot{\tilde{k}}_f = -\dot{\hat{k}}_f = -\eta_r \left\| r \right\|. \tag{27}$$

Proof: Define a Lyapunov function candidate as

$$L(r(t), \tilde{W}) = \frac{1}{2} r^{\mathrm{T}} M r + \frac{1}{2} \mathrm{Tr}(\tilde{W}^{\mathrm{T}} F^{-1} \tilde{W}) + \frac{\tilde{k}_f^2}{2\eta_r}. \tag{28}$$

Where F, η_r are the learning rate for the WFCMAC memory contents, error bound, respectively. By differentiating (28) with respect to time and using (26), (27), and using properties of the robot dynamics are introduced in Section 2, we can obtain.

$$\begin{aligned}
\dot{L} &= r^{\mathrm{T}} M\dot{r} + \tfrac{1}{2} r^{\mathrm{T}} \dot{M} r + \mathrm{Tr}(\dot{\tilde{W}}^{\mathrm{T}} F^{-1} \tilde{W}) + \tfrac{\tilde{k}_f \dot{\tilde{k}}_f}{\eta_r} \\
&= -r^{\mathrm{T}} K_{\mathrm{v}} r + \tfrac{1}{2} r^{\mathrm{T}} (\dot{M} - 2V) r + \mathrm{Tr}\tilde{W}(F^{-1}\dot{\tilde{W}}^{T} + b r^{\mathrm{T}}) + r^{\mathrm{T}}(\varepsilon - \hat{k}_f)\mathrm{sgn}(r) + \tfrac{\tilde{k}_f \dot{\tilde{k}}_f}{\eta_r} \\
&= -r^{\mathrm{T}} K_{\mathrm{v}} r + \tfrac{1}{2} r^{\mathrm{T}} (\dot{M} - 2V) r + \mathrm{Tr}\tilde{W}(F^{-1}\dot{\tilde{W}}^{T} + b r^{\mathrm{T}}) + \varepsilon\, r^{\mathrm{T}} - \hat{k}_f \left\| r \right\| - (k_f - \hat{k}_f) \left\| r \right\| \\
&= -r^{\mathrm{T}} K_{\mathrm{v}} r + \varepsilon\, r^{\mathrm{T}} - k_f \left\| r \right\| \\
&\leqslant -r^{\mathrm{T}} K_{\mathrm{v}} r + \left\| \varepsilon \right\| \left\| r \right\| - k_f \left\| r \right\| = -r^{\mathrm{T}} K_{\mathrm{v}} r - (k_f - \left\| \varepsilon \right\|) \left\| r \right\| \leqslant -r^{\mathrm{T}} K_{\mathrm{v}} r \leqslant 0.
\end{aligned} \tag{29}$$

Since $\dot{L}(r(t), \tilde{W}, \tilde{k}_f)$ is a negative semi-definite function, i.e. $\dot{L}(r(t), \tilde{W}(t), \tilde{k}_f(t)) \leqslant \dot{L}(r(0), \tilde{W}(0), \tilde{k}_f(0))$, it implies that r, $\tilde{W}$ and $\tilde{k}_f$ is bounded functions. Let function $h \equiv (k_f - \|\varepsilon\|)r \leqslant (k_f - \|\varepsilon\|)\|r\| \leqslant -\dot{L}(r, \tilde{W}, \tilde{k}_f)$ and integrate function $h(t)$ with respect to time

Because $\dot{L}(r(0), \tilde{W}(0), \tilde{k}_f(0))$ is a bounded function, and $\dot{L}(r(t), \tilde{W}(t), \tilde{k}_f(t))$ is a nonincreasing and bounded function, the following result can be achieved:

In addition, $\dot{h}(t)$ is bounded; thus, by Barbalat's lemma can be shown that $\lim_{t\to\infty} h(t) = 0$. It can imply that r will be converging to zero as time tends to infinity. □

5 Simulation and experimental results

5.1 Simulation results

A three-link deicing robot manipulator as shown in Fig.1 is utilized in this paper to verify the effectiveness of the proposed control scheme. The detailed system parameters of this robot manipulator are given as: link mass m_1, m_2, m_3 (kg), lengths l_1, l_2 (m) angular positions q_1, q_2 (rad) and displacement position d_3 (m).

The parameters for the equation of motion (1) can be represented as

$$\begin{cases} M(q) = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, V(\dot{q}) = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix}, \\ V_{11} = -8m_2 l_1 l_2 c_1 s_1 \dot{q}_1 + (-1/2 m_2 s_2 c_2 l_2^2 + m_3(-2s_2 c_2 l_2^2 - 2s_2 l_1 l_2)\dot{q}_2, \\ M_{11} = 9/4 m_1 l_1 + m(1/4 c_2 l_2 + l_1^2 + l_2 l_1 (c_1^2 - s_1^2)) + m_3(c_2 l_2^2 + l_2^2 + 2c_2 l_1 l_2), \\ M_{22} = 1/4 m_2 l_2^2 + m_3 l_2^2 + 4/3 m_1 l_1^2, M_{22} = 1/4 m_2 l_2^2 + m_3 l_2^2 + 4/3 m_1 l_1^2, \\ M_{33} = m_3, M_{12} = M_{13} = M_{21} = M_{31} = 0, \\ V_{21} = (-1/2 m_2 s_2 c_2 l_2^2 + m_3(-2s_2 c_2 l_2^2 - 2s_2 l_1 l_2)\dot{q}_1, V_{22} = -m_3 s_2 l_2 \dot{d}_3, \\ V_{23} = -2m_3 s_2 l_2 \dot{q}_2, V_{32} = -m_3 s_2 l_2 \dot{q}_2, V_{12} = V_{13} = V_{31} = V_{33} = 0, \\ G(q) = \begin{bmatrix} (1/2 c_1 c_2 l_2 + c_1 l_1) m_2 g \\ (-1/2 s_1 s_2 l_2 m_2 + c_2 l_2 m_3) g \\ m_3 g \end{bmatrix}. \end{cases} \tag{30}$$

Where $q \in \mathbb{R}^3$ and the shorthand notations $c_1 = \cos(q_1)$, $c_2 = \cos(q_2)$, and $c_1 = \sin(q_1)$, $c_2 = \sin(q_2)$ are used.

For convenience of simulation, nominal parameters of the robotic system are given as $m_1 = 3\,(\mathrm{kg})$, $m_2 = 2\,(\mathrm{kg})$, $m_3 = 2.5\,(\mathrm{kg})$, $l_1 = 0.14$ (m), $l_2 = 0.32$ (m) and and the initial conditions $q_1(0) = 1$, $q_2(0) = 0$, $d_3(0) = 0$, $\dot{q}_1(0) = 0$, $\dot{q}_2(0) = 0$, $\dot{d}_3(0) = 0$. The desired reference trajectories are $q_{d1}(t) = \sin(t)$, $q_{d2}(t) = \cos(t)$, $d_{d2}(t) = \cos(t)$, respectively.

For recording respective control performance, the mean-square-error of the position-tracking response is defined as:

$$\mathrm{MSE}_i = \frac{1}{T}\sum_{j=1}^{\mathrm{T}} [\, q_{di}(j) - q_i(j)\,]^2, \quad i = 1, 2, 3. \tag{31}$$

Where T is the total sampling instant, and q_i and q_{di} are the elements in the vector q_i and q_{di}. In this paper, the numerical simulation results carried out by using Matlab software.

For the proposed RASOWFCM control system, the parameters are chosen in the following:

$$\Lambda = \mathrm{diag}\,\{10, 10, 10\}\,, K_{\mathrm{v}} = \mathrm{diag}\,\{50, 50, 50\}\,, F = 15, \eta_r = 0.2. \tag{32}$$

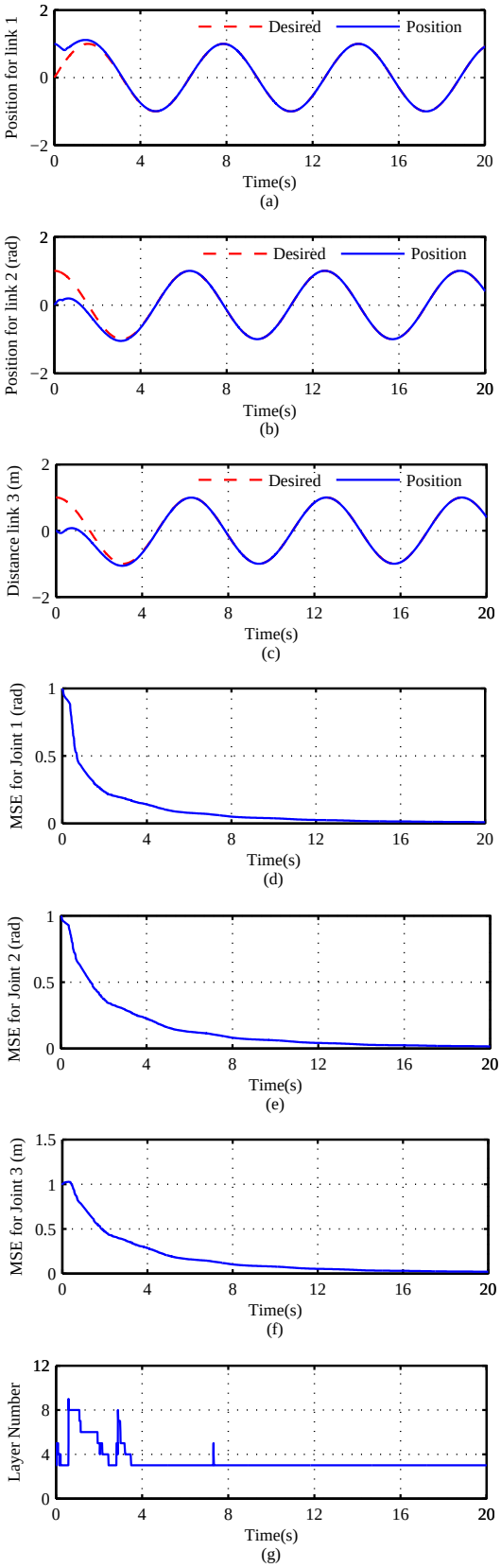


Figure 5: Simulated position responses, MSEs and layer number of the RA-SOWFCM control system at joints.

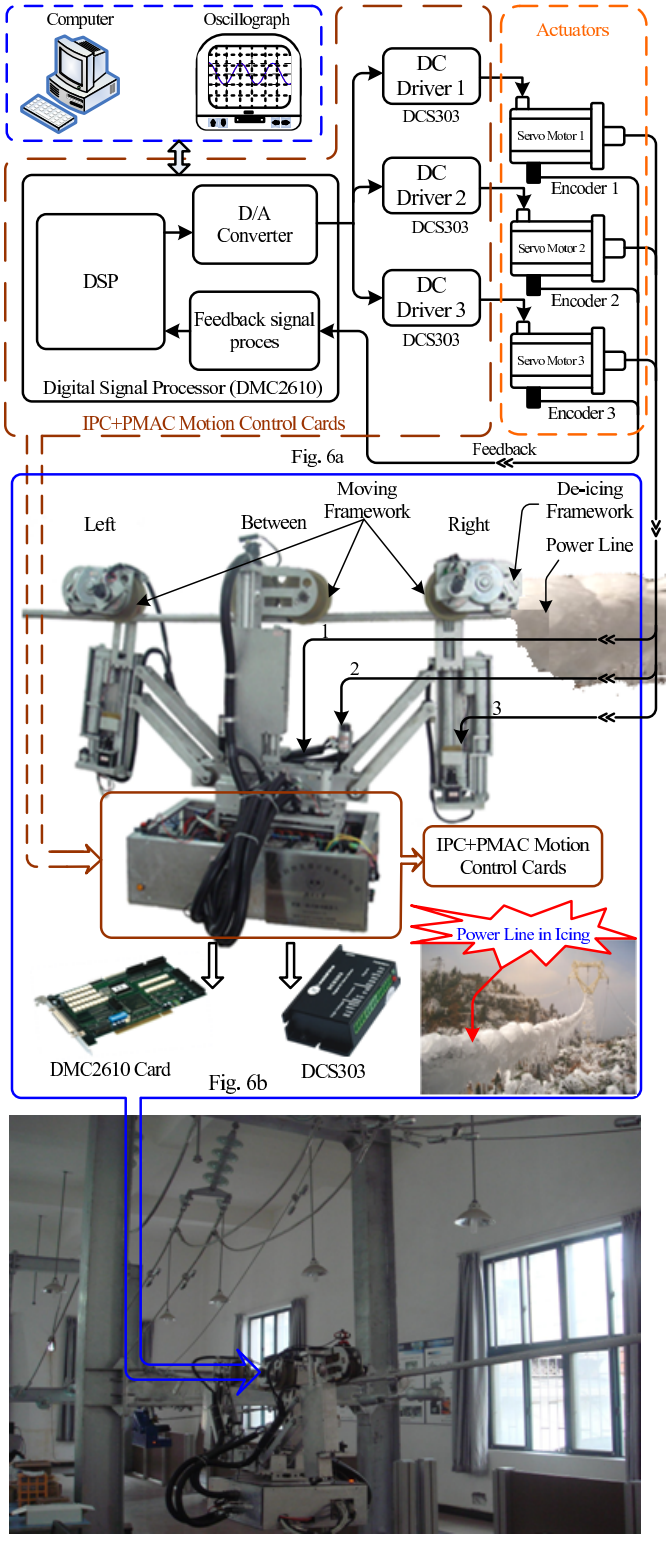


Figure 6: IIPC-based deicing robot position control system a) Block diagram of three-link deicing robot manipulator c) image of practical control system. control system, b) Image of practical control system.

and the initial values of system parameters are given as $n_k = 2$, $\hat{k}_f = 1$, the inputs of WFCMAC, the translation and dilation of wavelet functions are selected to cover the input space $\{[-1 \;\; 1][-1 \;\; 1][-1 \;\; 1]\}$. The threshold value of K_g is set as 0.1; K_c is set as 0.01. The simulation results of the proposed RASOWFCM system, the responses of joint position MSE and layer number are depicted in Fig.5(a), (b), (c); (d), (e), (f) and (g), respectively. The simulation results show that the proposed RASOWFCM control system can achieve a favorable tracking performance with self-organizing of WFCMAC, and the layers of WFCMAC will be converge to three layers.

5.2 Experimental results

An image of a practical experiment control system for deicing robot consists of three manipulators and is shown in Fig 6(b). The left and right manipulators have three links with two revolute joints and a prismatic joint. End-effectors of each manipulator have attached the motion structure to move the deicing robot on the power line and the deicing device. Under normal operating conditions, the left and right manipulators are only in operation. The middle manipulator has only two joints with a revolute joint and a prismatic joint. It only works when the deicing robot voids obstacles on the power line. In general, the operation of deicing robot is very complex. In this paper, we consider only the three-link deicing robot manipulator for proposed methodologies while the other manipulator is the same. The hardware block diagram of the control system is implemented to verify the effectiveness of the proposed methodologies and is shown in Fig.6(a). Each joint of manipulator is derived by the "EC**" type MAXON DC servo motors, and each motor contains an encoder. Digital filter and frequency multiplied by circuits are built into the encoder interface circuit to increase the precision of position feedback. The DCS303 is a digital DC servo driver developed with DSP to control the DC servo motor. The DCS303 is a micro-size brush DC servo drive. It is an ideal choice for this operating environment. Two DC servo motor motion control cards are installed in the industrial personal computer, in which, a 6-axis DC servo motion control card is used to control the joint motors and a 4-axis motion control card is used to control the drive motors. Each card includes multi-channels of digital/analog and encoder interface circuits. The model DMC2610 has a PCI interface connected to the IPC. The DMC2610 implements and excutes the proposed program in the real time. Considering that the control sampling rate $T_s = 1ms$ is too demanding for the hardware implementation, $T_s = 10ms$ is thus chosen here. The experimental parameters of the proposed RASOWFCM control system are selected as simulation: In this section, the control objective is to control each joint angle of a three-link deicing robot manipulator to move periodically for a periodic step commands and initial conditions of the system are given as $q_1\,(0) = 0\,(\text{rad})$, $q_2\,(0) = 0(\text{rad})$, $d_3\,(0) = 0\;(\text{m})$. Finally, the experimental position and tracking error response results of the proposed RASOWFCM control system are depicted in Fig.7 (a), (b) and (c). According to these experiment and simulation results of proposed RASOWFCM control system due to sinusoidal and periodic step reference commands indicate that the high-accuracy tracking position responses can be achieved by using the proposed RASOWFCM control system for difference reference commands under a wide range of external disturbance.

6 Conclusion

In this paper, a RASOWFCM control system is proposed to control the joint position of a three-link deicing robot manipulator. In the proposed RASOWFCM system, dynamical system is completely unknown in the control process. The adaptive tuning laws of WFCMAC parameters and error estimation are derived according to Lyapunov function so that a stability of the

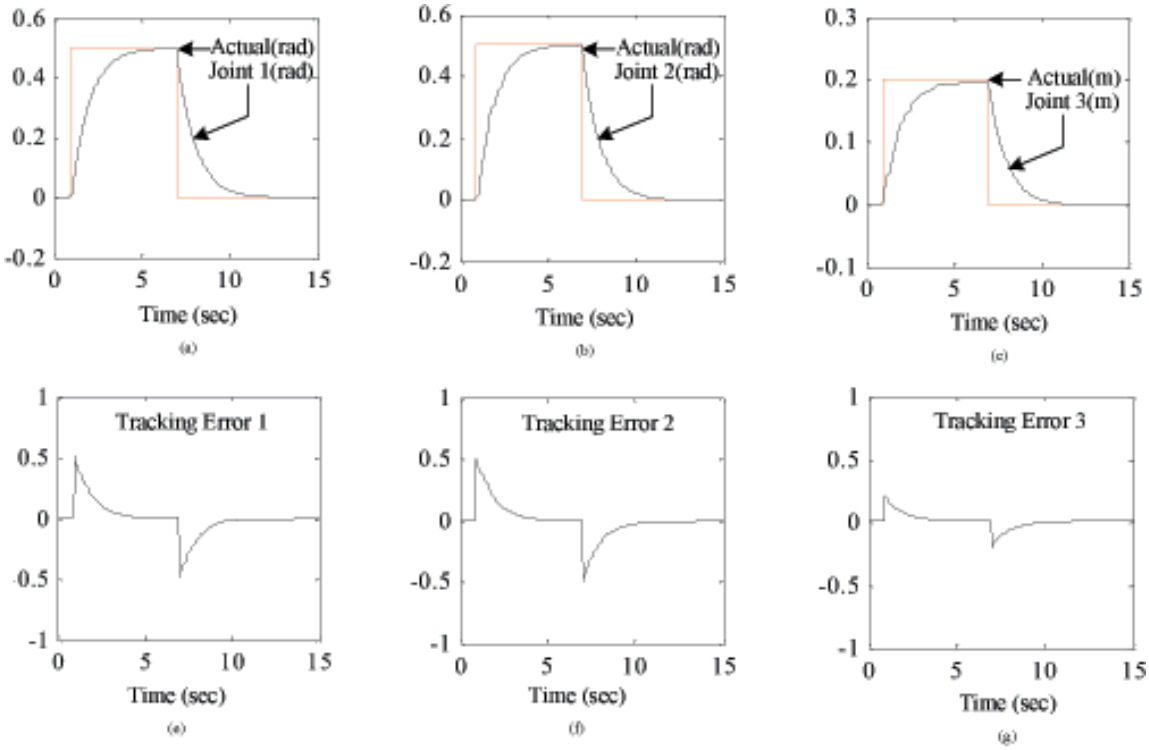


Figure 7: Experimental position responses, tracking errors of the proposed RASOWFCM control system at joints 1, 2 and 3

system can be guaranteed. This paper has not only successfully developed the RASOWFCM control system for a three-link deicing robot manipulator based on the novel WFCMAC requires low memory with online parameter tuning algorithm, but also provided novel architecture and mathematical model of deicing robot, which is verified the effectiveness through the practical application. Simulation and experimental results of the proposed RASOWFCM system can achieve favorable tracking performance for the proposed three-link deicing robot manipulator.

Bibliography

[1] M. Baban, C.F. Baban, C. Bungau, G. Dragomir, R.M. Pancu (2014); Estimation of the Technical State of Automotive Disc Brakes Using Fuzzy Logic, *International Journal of Computers Communications & Control*, 9(5): 531-538.

[2] C.R. Costea, H.M. Silaghi, D. Zmaranda, M.A. Silaghi (2015); Control System Architecture for a Cement Mill Based on Fuzzy Logic, *International Journal of Computers Communications & Control*, 10(2): 165-173.

[3] Y. Zou, Y. N. Wang, X. Z. Liu (2010); Neural network robust H_∞ tracking control strategy for robot manipulators. *Applied Mathematical Modelling*, 34(7): 1823-1838.

[4] Y. Feng, W. Yao-nan, Y. Yi-min (2012); Inverse Kinematics Solution for Robot Manipulator based on Neural Network under Joint Subspace, *International Journal of Computers Communications & Control*, 7(3): 459-472.

[5] C. Zhu, Y. F. Fang (2007); Adaptive control of parallel manipulators via fuzzy-neural network algorithm. *Journal Control Theory & Application*, 5(3): 295-300.

[6] T. Ngo, Y. Wang, T.L. Mai, M.H. Nguyen, J. Chen (2012); Robust Adaptive Neural-Fuzzy Network Tracking Control for Robot Manipulator. *International Journal of Computers Communications & Control*, 7(2): 341-352.

[7] C. F. Hsu, C. M. Lin, T. T. Lee (2006); Wavelet adaptive backstepping control for a class of nonlinear Systems, *IEEE Transcation Neural Network*, 17(5): 1175-1183.

[8] C. H. Lu (2009); Design and application of stable predictive controller using recurrent wavelet neural networks, *IEEE Transactions On Industrial Electronics*, 56(9): 733 – 3742.

[9] J. S. Albus (1975); A new approach to manipulator control: The cerebellar model articulation controller (CMAC), *J. Dyn. Syst. Meas. Control*, 97(3): 220 – 227.

[10] S. Jagannathan, S. Commuri, F. L. Lewis (1998); Feedback linearization using CMAC neural networks, *Automatica*, 34(3): 547 – 557.

[11] Y. H. Kim, F. L. Lewis (2000); Optimal design of CMAC neural-network controller for robot manipulators, *IEEE Transcation System Man Cybernation C, Application Revision*, 30(1): 22 – 31.

[12] C. T. Chiang, C. S. Lin (1996); CMAC with general basis functions, *Journal of Neural Network*, 9(7): 1199 – 1211.

[13] H. C. Lu, C. Y. Chuang, M. F. Yeh (2009); Design of hybrid adaptive CMAC with supervisory controller for a class of nonlinear system. *Neurocomputing*, 72(7-9): 1920 – 1933.

[14] Y. F. Peng, C. M. Lin (2004); Intelligent hybrid control for uncertain nonlinear systems using a recurrent cerebellar model articulation controller. *IEEE Proceedings Control Theory Application*, 151(5): 589 – 600.

[15] Y.C. Hsu, G. Chen, H.X. Li (2001); A Fuzzy adaptive Variable Structure Controller with Application to Robot Manipulators, *IEEE Transcation System Man Cybernation*, 31(3): 331 – 340.

[16] H. K. Khalil (1996); *Nonlinear systems*, Englewood Cliffs, NJ: Prentice-Hall, 1996.

[17] C. M. Lin, T. Y. Chen (2009); Self-organizing CMAC control for a class of MIMO uncertain nonlinear systems, *IEEE Neural Nets*, 20(9): 1377 – 1384.

[18] C.T. Lin, C. S. George Lee (1996); *Neural fuzzy systems*, Englewood Cliffs, NJ: Prentice-Hall.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):579-592, August, 2015.

A Trusted-based Cloud Computing Virtual Storage System and Key Technologies

K.H. Wu, L. Chen, Y. Li

Kehe Wu, Long Chen*, Yi Li
School of Control and Computer Engineering
North China Electric Power University
NO.2 Beinong Road, Changping District, Beijing 102206, China
lw_ncepu@163.com, easy_cl@163.com, somethingnew1989@163.com
*Corresponding author: easy_cl@163.com

Abstract: With the popularity of Cloud Computing, people become incresingly concern about security problems, especially the data security, which has become the biggest obstacle for the development of Cloud Computing. In order to protect confidentiality and integrity of user data in Cloud Computing, this paper firstly studies the relevant research works in fields of trusted computing and Cloud Computing data protection and secondly introduces the concept of trusted into Cloud Computing data protection, presents the concept of Trusted Virtual Block Storage Device (TVBSD) and designs the Trusted Cloud Computing Virtual Storage System (TCCVSS). And then, the key technologies such as isolation, block device encryption and two-way authentication are expounded in this paper. Finally, the result of experiments shows that the system and the related technologies can not only effectively ensure the security of user data, but also control the consequent performance overhead in a proper range.
Keywords: trusted, Cloud Computing, virtual storage, cloud storage, encryption, authentication.

1 Introduction

Cloud Computing is an extension of distributed computing and grid computing technology, which offers a variety of services to customers through the network, such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [1, 2]. Compared with the traditional PC service model, the Cloud Computing model provides almost limitless resources for users, which means inexhaustible, taken on demand and pay per use, and for the companies it can reduce the IT infrastructure investment and operation & support costs, so that the companies can gain the maximum economic profits.

Admittedly, cloud computing owns many merits, but it also has defects. As the computing tasks are performed in the cloud and all user data must be stored in the cloud, so the problems, such as where and how the data would be stored, are transparent to the user, which means that the user loses the absolute control over the data. Therefore, the service agreement signed between service providers becomes the only guarantee, and the confidentiality and integrity of data become the primary issues that should be considered [3, 4], and also become the biggest obstacle for the development of Cloud Computing.

In Cloud Computing there are two main data storage modes. One is distributed file system, which regard the virtual machine as a client for system storage the typical representative include GFS(Google File System) [5],Hadoop [6] and Amazon S3 [7]; and the other is virtual block storage device, the typical representative is Amazon EC2. Both have their advantages and disadvantages. The former, as its own application software must derectly minipulate user data, is difficult to guarantee the consistency of data; and the latter is more flexible, which is transparent, so that the customers can use it just as they usually use local disk, and thus the users can build their own

file systems and databases according to different needs. It not only can meet the requirements of massive digital information storage and improve the utilization of storage space, but also can shield the heterogeneity of operating system to realize storage management of automation and intelligentce, so as to better improve the quality of Cloud Computing services. Therefore, it is very significant to provide a safe and reliable data storage system and protection technologies for Cloud Computing.

Many existing data storage system has good reliability, flexibility, efficiency and ease of management performance, but it still need to be improved for ensuring data security. Though the exsiting storage system has used protecting strategies and technologies, including encryption, firewalls and rights management, it only has solved part of the security problems, which has a degree of limitation [8]. So there are still existing following problems in terms of ensuring the data safety and credibility:

(i) The problems of internal attack. using the traditional technologies such as encryption and firewalls can only defend the illegal intrusion from outer, but it is incapable to defend the internal attack. In the cloud computing model, the cloud services administrator (simply administrator for short) has many authority, if the insiders obtain the administrator privileges, he can attack the system from inside by an "legal" method, and then destroy the user data and the system. The administrator has all control permissions on the data, services, network and other important resources, so it can access to user data at any time or change the configuration file of the storage devices. Therefore, we must balance distribution of authority between the ordinary users and administrator or restrict administrators access to user data. Only in this way can we solve the problems of internal attack, and ensure the confidentiality of user data.

(ii) The problems of device management. As the characteristics of Cloud Computing, all user data are stored in the cloud, and in order to protect the confidentiality of the data, it will be stored with encrypted mode and be decrypted when using. However, as the storage device has no protection, the intruders can attack the device to obtain any data they want, and thus use attack technique the brute force attack to get the data that available to them, resulting in leakage of user data and destruction the confidentiality and integrity of data. Therefore, there is a must to practice rational management to storage device, for increasing protection measure of devices, which means to prevent attacks to the devices, so as to better prevent user data from unauthorized intrusion and tampering.

(iii) The problems of secure authentication. The following two situations can lead to the disclosure of user data: one is the storage device is remapped to the illegal users or accessed directly by the administrator; and the other is legitimate users stored their confidential data to a non-credible (illegal) storage device. In order to prevent the above situations from occurring, we must provide security authentication, through which, we can bind the storage device with legitimate user, making the legitimate users can not store data on a non-certified storage device, while the trusted storage devices can not be accessed by the unauthenticated user, that means user can only access the storage device belonging to its own. Only in this way can we completely prevent the above situations from occurring and thus can prevent data leakage.

In order to solve the above problems, this paper presents a trusted cloud computing virtual storage system based on the research of trusted computing and virtual storage technology, and the key techniques is researched and achieved. The main contributions are shown as follows:

(i) Through the study and research of trusted computing [9] and virtual storage technologies, this paper presents the concept of trusted virtual block storage devices, and then design a trusted cloud computing virtual storage systems , so as to better ensure the confidentiality and the integrity of data.

(ii) As for the problems of internal attack, this paper proposes the isolation technology, which separate the management section in user-level and the executive section in kernel of the storage

system, and then place them in different virtual machine, so that even the administrator cannot access the data directorly, only through the interfaces provided by the storage system can he manage the trusted virtual block storage device Thus, in this way it can solve the data security problems caused by administrator's overmuch operating authorities.

(iii) As for the problems of device management, this paper puts forward the block device encryption technology, which not only can encrypt the data stored in the trusted virtual block storage device, but also can encrypt the block storage device, so that only the user with the private key of the block storage device can access the device and obtain the data, while the unauthorized user cannot access it. In this way, it can avoid the illegal users obtain user data through attacking the storage devices, and thus well-ensured the security of the data.

(iv) As for the problems of secure authentication, this paper raises a two-way authentication technology, namely the user needs to authenticate the storage device whether it is a trusted one, and at the same time, the device also needs to authenticate the user whether it has access, and thus combine the user and device, making the legitimate user can only access the device belongs to its own. That is, the legitimate users can not store data on a non-certified storage device, and the trusted storage devices can not be accessed by the unauthenticated user.

(v) The effectiveness and performance of the system has been tested and analyzed, the result shows that the system and the related technologies not only can effectively ensure the security of user data, but also can control the consequent performance overhead in a proper range.

The paper is organized as follows. In section 1, we study the relevant research works in fields of trusted computing and cloud computing data protection. In section 2, we introduce the concept of trusted into cloud computing virtual storage, propose the concept of Trusted Virtual Block Storage Device (TVBSD) and design the Trusted Cloud Computing Virtual Storage System (TCCVSS), and then provide highlights of the key technologies, such as isolation, block device encryption and two-way authentication. In section 3 , we conduct test and analyze the effectiveness and performance of the system. The last section is the conclusion of the paper and gives lights to our future work.

2 Related Work

In recent years, as cloud computing technology gradually become mature, there are more and more studies on it., among which the data security become one of the crucial researches.

Some studies focus on the integrity of software load time and run time, enhancing system security by software integrity verification, so as to ensure the data security. For instance, the Seshadri [10] through memory virtualization technology can ensure that only validated code can execute in kernel mode, so as to withstand the attack to the integrity of the kemel caused by code injection and ensure system security. Xu et al. [11] propose a method based on the virtual machine monitor to detect and prevent the integrity of kemel, and hence prevent system from illeagal operating intrusion.

In addition, some studies mainly through the key management to protect the security of data. For example, Pearson et al. [12] use the anti-attack ability of TPM(Trusted Platform Module) and the hardware encryption technology to protect data encryption key, and combine the key with platform, making the key incapable to be used on the other platforms ,so as to protect the confidentiality of the data. Wang et al. [13] propose a management approach to key-used times based on trusted platform module in cloud storage, it can also control the times of using key while ensuring the safe storage of key, making the use of the key in user client controlled. Cheng et al. [14] present a novel key management scheme based on global logical hierarchical graph (GLHG), which is used to enforce correctly the global authorization policies of all users.

It can eliminates the redundant to minimize the amount of keys transferred and stored, and to maximum reduce the costs and risks of user key management.

As the trusted technology widely used in data security and platform security, there are also some research work applied the trusted technology to cloud computing to solve the credibility problem in virtual storage. Liu et al. [15] propose a new storage process of object-oriented properties in trusted storage, which can verify the QoS attributes of storage to resist a potential threat to the security and integrity of storage. However, this process only solve the trusted storage problem of physical storage device in single platform. Wang et al. [16] present a method to solve the security of data storage in virtualization platform, which is a data encapsulation method based on the properties of component and achieved by using the Trusted Platform Module (TPM). Yang et al. [17] design an architecture of cloud storage system based on TPM. This architecture used the symmetric key to encrypt data, and then used the asymmetric key to encrypt symmetric key, and finally used TPM to protect the asymmetric key. Only in this way can we manage the process of key storage, backup and share effectively.

Based on the above studies and inspirations, this paper introduce the trusted technology into the virtual storage of cloud computing, design a trusted cloud computing virtual storage system (TCCVSS), and put forward the isolation, block device encryption and two-way authentication technologies. Meanwhile, by testing and analyzing its effectiveness and performance of the system, the result shows that the system and the related technologies can effectively ensure the security of user data. The following sections describe the system architecture and key technologies in detail.

3 TCCVSS and Key Technologies

3.1 The Architecture of TCCVSS

Based on the research of trusted and storage technologies, this paper designs a Trusted Cloud Computing Virtual Storage System (TCCVSS) applied to IaaS. Each trusted virtual block storage devices (TVBSD) of the system are built on the storage resources based on hard disk and iSCSI, which can be used by multiple users but only one owner. TVBSD is a virtualization of the trusted physical storage device, which its storage space can be divided into two parts: one is a private storage space, which mainly used to store the confidential information of TVBSD, such as the key of TVBSD, the integrity measurement value, the access control list and so on; and the other is a user storage space, which mainly used to store the user data. TCCVSS is a trusted virtual storage system based on the virtualization platform framework of Xen, which inherits the functions in traditional storage system including snapshot and mirroring, and also joins the isolation, block device encryption and two-way authentication technologies, making the protection of system data more safe and reliable. Fig.1 shows the system architecture diagram of TCCVSS.

Among these, the main components are listed as follows:

(1) The TVBSD Manager Tool: Implementation of this section is based on Logical Volume Manager (LVM2), and it is mainly used to manage the logical volume, such as creating, deleting, snapshotting, mirroring and dynamic extension capabilities.

(2) The TVBSD Master: This section can be achieved on the basis of Device Mapper, which is the new component of Linux2.6 kernel and the later versions. Device Mapper provides a method to unify the ways for creating the virtual layer of block device, so it can facilitate realizes the functions including striping, tandem, mirroring and snapshots in virtual layer for users. This paper use the isolation technology, which will be discussed in later chapters, to separate this section into two parts: TVBSD Master front-end and TVBSD Master back-end, so that it can

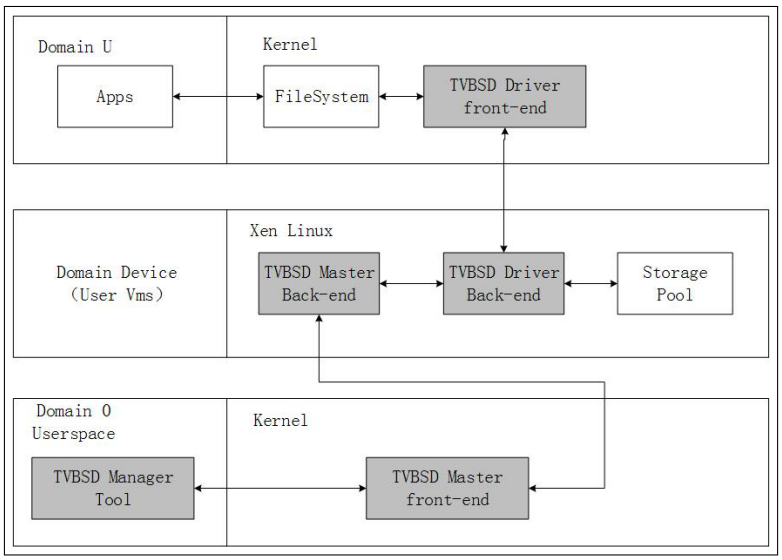


Figure 1: System Architecture Diagram of TCCVSS

effectively prevent the illegal access to data.

(3) The TVBSD Driver: This section mainly provides system security functions including access control, data encryption and user authentication, which can guarantee the confidentiality and integrity of data. The encryption and user authentication technologies will be discussed in more detail in later chapters.

3.2 Isolation Technology

In the traditional cloud computing model, the administrator has so many authority that if the insiders obtain the administrator privileges, it can attack the system from within, and then destroy the user data and the system. The administrator has all control permissions to the data, services, network and other important resources, so it can access to user data at any time and change the configuration file of the storage devices. Therefore, we must balance distribution of authority between the ordinary users and administrator or restrict administrators access to user data. Only then can we solve the problems of internal attack, and ensure the confidentiality of user data.

In the traditional virtual storage system, such as in Xen, the virtual block device (VBD) and its management tool are in the same operating system, as shown in Fig.2 by dotted box. In this way, the administrator can access the VBD directly through VBD Master Tool while the system is running, and then illegal intrude or tamper with the data stored in VBD to destroy the confidentiality and integrity of data.

In order to protect the security of user data, there is a need to prevent administrator to access the user data directly, so this paper present the isolation technology to solve this problem. That is, to divide the VBD Master into VBD Master front-end and VBD Master back-end, as shown in Fig.3. And then place the VBD Master front-end and VBD Master Tool in the same virtual machine to provide VBD management interface for VBD Master Tool; and place the VBD Master back-end in another virtual machine to receive the orders from VBD Master front-end and execute them, as shown in Fig.4 by dotted box. In this way, the executing section of VBD is separated from VBD Master Tool, so the administrator does not have permission to access the virtual machine where the VBD Master back-end resides. Therefore, the administrator can only access VBD through the interface provided by VBD Master front-end instead of accessing

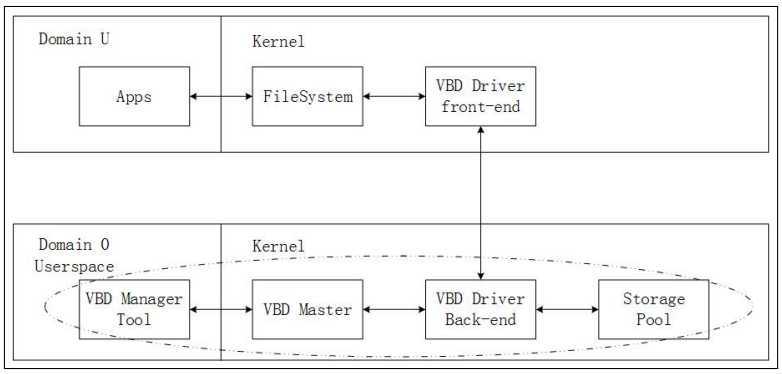


Figure 2: The Organization Chart of VBD in Xen

directly, thus ensuring the confidentiality and integrity of data.

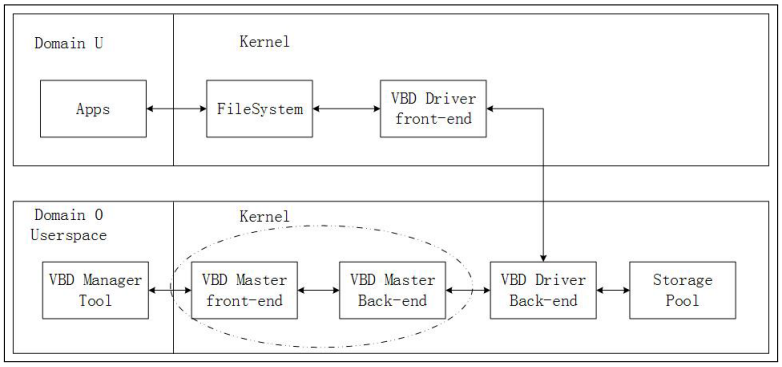


Figure 3: VBD Master Divide Into VBD Master Front-end and VBD Master Back-end

3.3 Block Device Encryption Technology

Since the characteristics of Cloud Computing, all user data is stored in the cloud, and in order to protect the confidentiality of the data, it will be stored with encrypted mode and be decrypted when using. However, due to the storage device does not make any protection, the intruders can attack the device to obtain any data they want, resulting in leakage of user data and destruction of data confidentiality and integrity. Therefore, there is a need for rational management of storage devices, to increase protection measure of devices, which means to prevent attacks on the devices, so as to better protect user data from unauthorized intrusion and tampering.

To this end, this paper presents a block device encryption technology, which means assigning a Block Device Encryption Key (BDEK) for each trusted virtual block storage device(TVBSD). During the initializing phase, the system generates a BDEK for each user, which is an asymmetric key consisted of two keys–a public key and a private key. The users own the private key themselves to protect the key from exposure to any other entity; and when the user need to store data, the system creates TVBSD according to user needs, and then the public key certificate issued to each TVBSD. The relationship between user and TVBSD is shown in Fig.5 as follows:

It can be seen from the figure that there is a one-to-many relationship between the user and the TVBSD, that is each user can have multiple TVBSD and each TVBSD corresponds to only one user. And meanwhile, each TVBSD has just one Block Device Encryption Key (BDEK), which is used to sign the data information rather than encrypt it. If we used K_{pri} and K_{pub} to

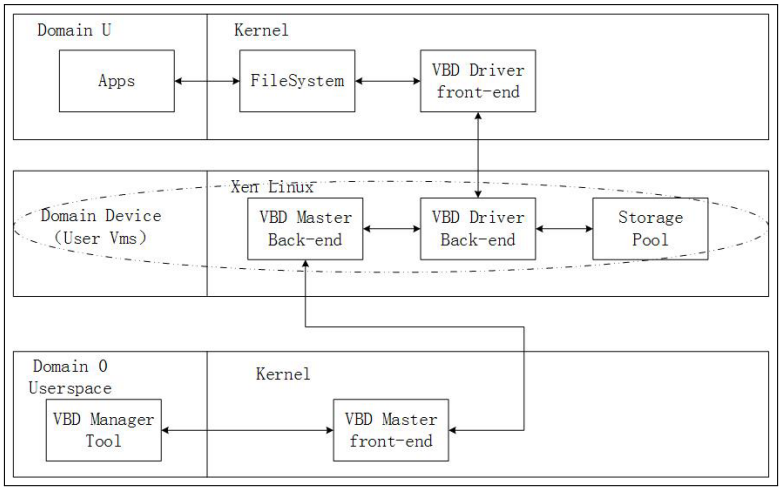


Figure 4: The Isolation Technology in TCCVSS

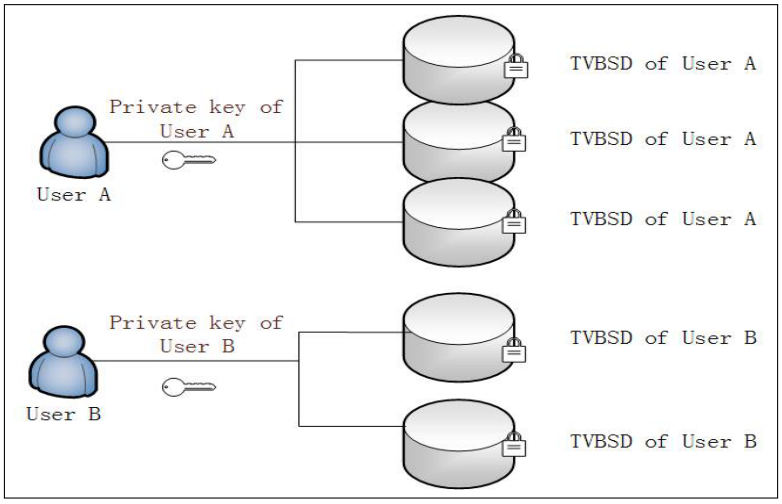


Figure 5: The Relationship Between TVBSD and User

represent private key and public key of BDEK, and used the Data to represent the ciphertext, then the data read and write operations using BDEK as shown below:

(1) Data write operation: use the public key of BDEK to sign the data, which is expressed as follows:

$$Data_1 = E(Data, K_{pub}) \tag{1}$$

where the $E(Data, K_{pub})$ means using the public key K_{pub} to sign the data $Data$, the $Data$ means the ciphertext and the $Data_1$ means the actual stored information in the block device, which is signed using the public key, and only the one who has the corresponding private key can decrypt the ciphertext to get the useful information.

(2) Data read operation: we need to isolate the information $Data$ from the signed data, which is expressed as follows:

$$Data = D(Data_1, K_{pri}) \tag{2}$$

where the $D(Data_1, K_{pri})$ means using the private key to decrypt the signed data $Data_1 = E(Data, K_{pub})$, and the $Data$ means the ciphertext. Thus it can be seen that only the one who has the corresponding private key can decrypt the data signed by the public key, so as to protect the data from obtaining by the unauthorized entities.

In summary, using the BDEK technology can prevent the unauthorized entities from accessing the TVBSD, even under the condition of the device be attacked or lost, so as to ensure the security and credibility of TVBSD and Protect the data from being invaded and abused.

3.4 Two-way Authentication Technology

The following two situations can lead to the disclosure of user data: one is the storage device is remapped to the illegal users or accessed directly by the administrator; and the other is legitimate users stored their confidential data to a non-credible (illegal) storage device. In order to prevent the above situations from occurring, we must provide security authentication, through which, we can bind the storage device with legitimate user, making the legitimate users can not store data on a non-certified storage device, while the trusted storage devices can not be accessed by the unauthenticated user, that means user can only access the storage device belonging to its own. This paper proposed the two-way authentication technology to solve the above problems, which is mainly shown in the following two aspects:

(1) Can block the storage requests initiated by the user who does not have the access to trusted virtual block storage device.

(2) Can reject the unauthorized virtual block storage device as users storage device.

In order to better complete the two-way authentication process, this paper proposed a Trusted Authentication Module (TAM) and designed three tables, namely: the user information table User_list, the trusted virtual block storage device table Tvbsd_list and the corresponding relation table of user and TVBSD is User_Tvbsd_list. Among this, the User_list is used to store the information of all authenticated users, while the Tvbsd_list is used to store the information of all trusted devices, and the User_Tvbsd_list is used to store the mapping relationship between the users and their corresponding TVBSD. The specific authentication steps are as follows:

The process of two-way authentication is shown in the following Fig.6:

(1) The user send a data storage request to TAM, including the information of user information ID_{user}, device information ID_{tvbsd} and so on.

(2) When the TAM received the data storage request:

(a) To search the user information table User_list based on the user information ID_{user}, determining whether the user is an authenticated one. If so, then go to next step; if not, then reject the user's data storage request and feedback to the user.

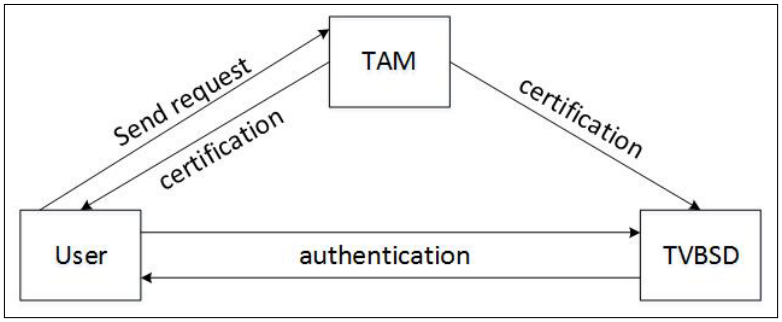


Figure 6: The Two-way Authentication

(b) To search the TVBSD table Tvbsd_list based on the device information ID_{tvbsd}, determining whether the device is an authenticated one. If so, then go to next step; if not, then reject the user's data storage request and feedback to the user.

(c) If the user and the device are both authenticated, then continue to search the corresponding relation table of user and TVBSD User_Tvbsd_list, determining whether the block storage devices users applied to belongs to users themselves. If so, then allow the data storage request and issue the certificate to the user and the TVBSD user applied to; if not, then reject the user's data storage request and feedback to the user.

(3) To establish a connection between user and TVBSD to complete the two-way authentication:

(a) User creates a timestamp T_1 and then sent it with user's certificate to the TVBDS.

(b) TVBDS creates another timestamp T_2 and calculate

$$M_1 = E(T_2||H(Data)||E(ID_{user}||T_1||T_2, K_{pubU}), K_{priT}) \tag{3}$$

then sent its certificate with timestamp T_2 and M_1 to the user. Among this, $H(Data)$ means the hash value of information in this communication, K_{pubU} means the user's public key, K_{priT} means the TVBDS's private key, and $E(Data, K)$ means encrypting the information $Data$ with the key K.

(c) To verify the TVBSD. Firstly, to decrypt the M_1 with the public key of TVBSD K_{pubT} to get the information including T_2, $H(Data)$ and $E(ID_{user}||T_1||T_2, K_{pubU})$. And then, to decrypt the $E(ID_{user}||T_1||T_2, K_{pubU})$ with private key of user K_{priU} to get the information including ID_{user} and T_1, determining whether the information we got matched the original one. If so, the user accept the TVBSD; if not, the user reject the TVBSD.

$$Verity(ID_{user}||T_1||T_2, M_1) = True \tag{4}$$

(d) The user calculate

$$M_2 = E(H(Data)||E(ID_{tvbsd}||T_2, K_{priU}), K_{pubT}) \tag{5}$$

and then sent it to TVBSD.

(e) To verity the user. Firstly, to decrypt the M_2 with the private key of TVBSD K_{priT} to get the information including $H(Data)$ and $E(ID_{tvbsd}||T_2, K_{priU})$. And then, to decrypt the $E(ID_{tvbsd}||T_2, K_{priU})$ with public key of user K_{pubU} to get the information including ID_{tvbsd} and T_2, determining whether the information we got matched the original one. If so, the TVBSD accept the user; if not, the TVBSD reject the user.

$$Verity(ID_{tvbsd}||T_2, M_2) = True \tag{6}$$

The whole procedure of two-way authentication is finished after the above steps, which can implement the binding between user and TVBSD, so as to better ensure the data confidentiality and integrity.

4 Experiments and Analysis

4.1 Validity Analysis

In this paper, we designed four experiments to evaluate the effectiveness of the system, from the perspective of the kind of attack, it can be divided into the following tests: unauthorized user access attack test, illegal equipment mount attack test, data monitoring attack test and physical device attack test.

(1) For the unauthorized user access attack test: the access of unauthorized user can be sorted into the following two situations: one is the access of illegal user without authorization; and the other is the access to the storage device, which does not belong to the authenticated user. However, in traditional virtual storage systems, take Xen for example, as administrators have too much operating authority, so both of the above users can use the administrator privileges to access to the users storage device directly, or remap the storage device to the unauthorized users, so as to obtain the users' data. While in the TCCVSS proposed in this paper, owning to the use of isolation technology, the administrator can only use the interface provided by the TVBSD Master front-end to access to the TVBSD, which prevent the administrator from accessing to the virtual block storage devices directly. Meanwhile, the Trusted Authentication Module (TAM) in two-way authentication technology can verify the identity of a user when the user initiate the request for accessing. The experiments result indicates that only the users, who can pass the validation of TAM, can establish a connection with the TVBSD, while the illegal users, who can not pass the validation of TAM, can not access to the TVBSD, so that it ensure the security of user data.

(2) For the illegal equipment mount attack test: Similar to the unauthorized user access, the access of illegal device mounts can also be divided into the following two situations: the first is the mount of equipment without authentication; the second is the authorized device mounted to the user who does not have the access to the device. In the TCCVSS, the Trusted Authentication Module (TAM) in two-way authentication technology can verify the identity of a user when user initiate request for mounting. The experiment suggests that only authenticated devices can establish a connection with user, while the illegal device would be denied by TAM, and thus ensure the security of user data.

(3) For the data monitoring attack test: In traditional storage systems, VBD management tools and VBD are in the same system, the system administrator can monitor the device's memory by installing and running various software, and can steal user data by attacking the memory. In the TCCVSS, with the use of isolation technology, the management section of TVBSD and the actual operation section of TVBSD can be isolated, so that the administrator can only access to the TVBSD through the interface provided by the TVBSD Master front-end. The result of testing indicates that the administrator is unable to install monitoring software in the equipment area where the perform section of TVBSD stayed, thereby, he cannot carry out monitoring attack to the device's memory, and thus ensure the security of the data.

(4) For the physical device attack test: In the traditional storage systems, both the attack on a physical storage device and the loss of the physical storage device may cause damage to user data. In the TCCVSS, with the use of block device encryption technology, the security of the device can be ensured. After testing, the results shows that it is difficult to steal user data by attacking the physical device in the absence of private key, and thus ensure the security of data.

The above experiments indicates that the system can perfectly ensure the confidentiality and integrity of data by using various technologies, such as the isolation, block device encryption and tw-way authentication, etc. Compared with the traditional virtual storage system, the TCCVSS has a higher level of security and reliability.

4.2 Performance Analysis

The experimental environment in this paper is as follows:
CPU: Inter Core 2 @2.7 GHz
RAM: 2GB
System: CentOS 6.4, the kernel version is 2.6.32
Platform: Xen v4.2.2
Hash function: SHA-1
Encryption algorithm: AES in OpenSSL0.9.8
Test tools: IOMeter, IOzone

In order to test the performance of the system designed in this paper, we use the test tools IOMeter and IOzone to measure the effects of this system on the performance of Xen virtualization platform.

Iometer is an I/O subsystem measurement and characterization tool for single and clustered systems. Its main functions include testing the performance of the disk and the network controllers, bandwidth and delay capacity of the bus, the throughput of the network, the performance of the shared bus, the performance of the hard drive in system level and the performance of the network in system level, etc [18]. Therefore, it can be used as a benchmark and troubleshooting tool and is easily configured to replicate the behavior of many popular applications.

IOzone is a file system benchmark tool, which can test the read/write performance of file system in different operating system. The benchmark generates and measures a variety of file operations,such as Read, write, re-read, re-write, read backwards, read strided, fread, fwrite, random read, pread, mmap, aio_read, aio_write,etc. Iozone has been ported to many machines and runs under many operating systems.

In order to better measure the performance of this system designed in this paper, we access the following three types of disk in Xen: the disk with eCryptfs, the disk with dm-crypt and the disk with TVBSD. The eCryptfs is the enterprise file encryption system in Linux; And the dm-crypt is a target driver developed based on the Device-Mapper, which mainly offers transparent encryption of block devices.

First, we use IOMeter tool to do three experiments under the same operating environment, to compare the throughput of the disk under three cases. The test parameters set as follows:

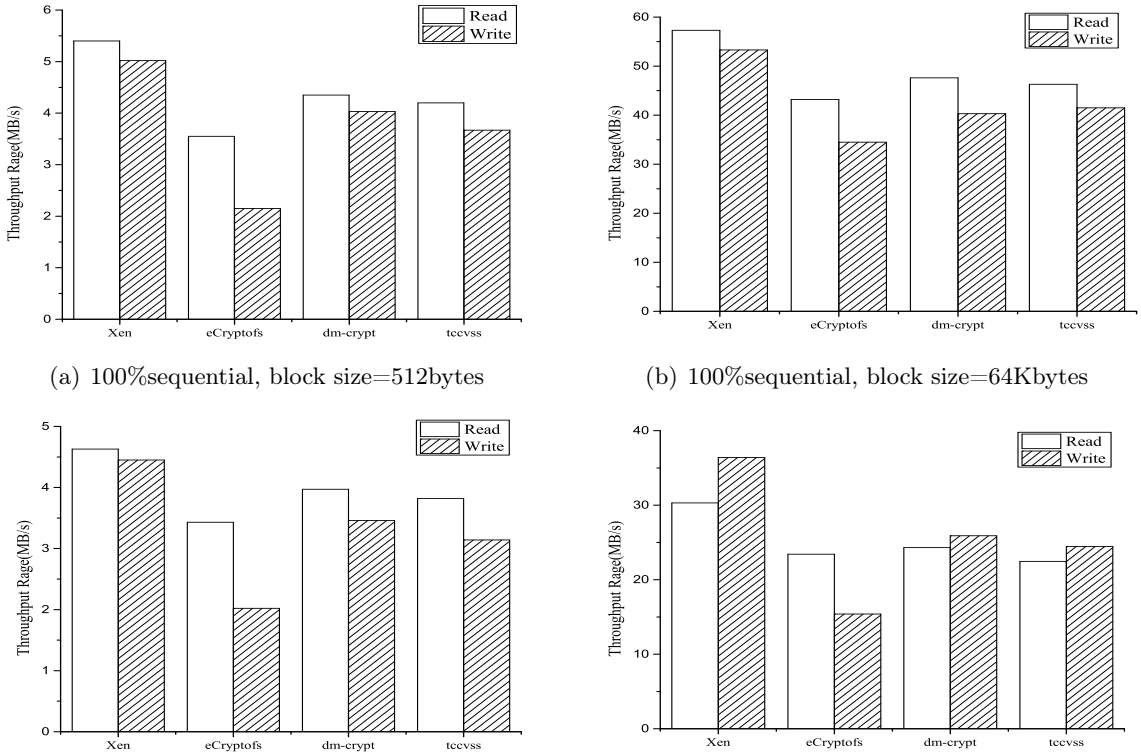
(1) The size of data block: transfer request size is set to two cases: 512bytes and 64Kbytes;

(2) The method of read and write: percent random/sequential distribution is set to two cases: 100% sequential and 100% random;

(3) The percentage of read and write: percent read/write distribution is set to two cases: 100% read and 100% write .

The result is shown in Fig.7, which is the average of multiple measurement experiments:

It can be seen that, compared with Xen system without encryption function, the read and write performance of the other three disks has a certain loss, which is caused by adding the encryption function, and thus increase the system overhead. Compared with the disk of dm-crypt, the TCCVSS only has a small portion of performance lost, which is not obvious. But compared with the disk of eCryptfs, the performance improved significantly. Meanwhile, in the case of the same block size, the throughput of sequential read and write is higher than that of random read and write.



(a) 100%sequential, block size=512bytes

(b) 100%sequential, block size=64Kbytes

(c) 100%random, block size=512bytes

(d) 100%random, block size=64Kbytes

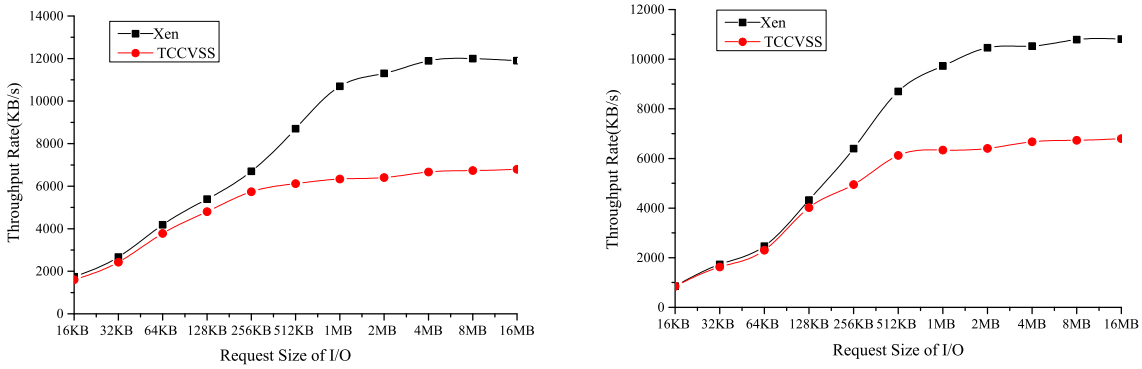
Figure 7: Throughout Evaluation

Then we use IOzone tool to measure its I/O performance being a file system server. The experiment focus on comparing the change of throughout rate of the original Xen system with that of this system under the circumstance of different I/O request size, the requested block size change from 16KB to 16MB. The results show in Fig.8:

As is shown in this Figure, with the increase of I/O requests, the system's performance gap also increase. This is because with the increase in I/O request, the time system spent in locating the file location and reading the document properties increases small, but the proportion of encryption and certification in the entire operation increases because of joining some technologies, such as the isolation, block device encryption and two-way authentication, so the CPU processing speed of encryption and authentication operations become the bottleneck of system performance, which is not obvious when the I/O request is very small.

5 Conclusion

Based on the research of the trusted computing and storage technologies, this paper introduces the concept of trusted into cloud computing virtual storage, defines the concept of Trusted Virtual Block Storage Device (TVBSD) and designs the Trusted Cloud Computing Virtual Storage System (TCCVSS), and then provides highlights of the key technologies, such as isolation, block device encryption and two-way authentication, with which can we ensure the safety of the user data in the system. We use the simulation experiments to test its effectiveness, the results show that the system can well protect the confidentiality and integrity of the user data. And



(a) The performance comparison of write operations (b) The performance comparison of read operations

Figure 8: Performance Comparison Between Xen and TCCVSS at Different I/O Request Size

the test tools, such as IOMeter and IOzone have been used to test its performance, the results show that this system, compared with other storage systems, has improved to some extent in protecting the confidentiality and the integrity of user data. However, the system loss increased due to use the technologies, such as isolation, block device encryption and two-way authentication, resulting in a performance degradation within a reasonable range. In addition, because the system is designed in an ideal experiment environment, there are many problems need to be solved in the practical application, so the further work is to improve the performance of the system under the condition of ensuring the confidentiality and integrity of user data, and to implement a more safe and reliable cloud computing virtual storage system, which can be suitable for practical application scenes.

Bibliography

[1] B. Tograph, Y.R. Morgens (2008), Cloud computing, *Communications of the ACM*, 51(7): 9-11.

[2] A. Weiss (2007), Computing in the clouds, *Network of ACM*, 11(4): 16-25.

[3] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica (2009), Above the clouds: A Berkeley view of cloud computing. *Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS 28, 13. 2009.*

[4] J. Heiser, M. Nicolett (2008), Assessing the security risks of cloud computing. *Gartner Report.*

[5] http://labs.google.com/papers/gfs.html.

[6] http://hadoop.apache.org/.

[7] http://aws.amazon.com/s3/.

[8] U. Kűhn, K. Kursawe, S. Lucks, A.R. Sadeghi, C. Stűble (2005), Secure data management in trusted computing, *In: Cryptographic Hardware and Embedded SystemsCHES 2005: Springer*, 324-338.

[9] C. Shen, H. Zhang, H. Wang, J. Wang, B. Zhao, F. Yan, F. Yu, L. Zhang, M. Xu (2010), Research on trusted computing and its development. *Science China Information Sciences*, 53(3): 405-433.

[10] A. Seshadri, M. Luk, N. Qu, A. Perrig (2007), SecVisor: A tiny hypervisor to provide lifetime kernel code integrity for commodity OSes, *ACM SIGOPS Operating Systems Review*, 41(3): 335-350.

[11] M. Xu, X. Jiang, R. Sandhu, X. Zhang (2007), Towards a VMM-based usage control framework for OS kernel integrity protection. *In: Proceedings of the 12th ACM symposium on Access control models and technologies: ACM*, 71-80.

[12] S. Pearson, Y. Shen, M. Mowbray (2009), A privacy manager for cloud computing. *In: Cloud Computing: Springer*, 90-106.

[13] L. Wang, Z. Ren, Y. Dong, R. Yu, R. Deng (2013), A management approach to key-used times based on trusted platform module in cloud storage. *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, 50(8): 1628-1636.

[14] F. Cheng, Z. Peng, W. Song, S. Wang, Y. Cui (2013), Key management for access control in trusted cloud storages, *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*,50(8): 1613-1627.

[15] L. Zhaobin, Q. Wenyu, L. Keqiu, F. Ruoyu (2009), Object oriented property attestation for trusted storage. *In: IEEE 9th International Conference on Computer and Information Technology, CIT 2009, October 11, 2009 - October 14, 2009 Xiamen, China: IEEE Computer Society*, 93-97.

[16] D. Wang, D. Feng (2010), A hypervisor-based secure storage scheme. *In: 2nd International Conference on Networks Security, Wireless Communications and Trusted Computing, NSWCTC 2010, April 24, 2010 - April 25, 2010 Wuhan, Hubei, China: IEEE Computer Society*, 81-86.

[17] X. Yang, Q. Shen, Y. Yang, S. Qing (2011), A way of key management in cloud storage based on trusted computing. *In: 8th IFIP International Conference on Network and Parallel Computing, NPC 2011, October 21, 2011 - October 23, 2011 Changsha, China: Springer Verlag*, 135-145.

[18] J. Chen (2011), Design and Implementation Volume-Based Hierarchical Storage System. *Huazhong University of Science & Technology.*

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(4):593-606, August, 2015.

Automated Test Sequence Optimization Based on the Maze Algorithm and Ant Colony Algorithm

W. Zheng, N.W. Hu

Wei Zheng*, Naiwen Hu
National Engineering Research Center of
Rail Transportation Operation and Control System
Beijing Jiaotong University
Beijing 100044, China
wzheng1@bjtu.du.cn, 12120318@bjtu.edu.cn
*Corresponding author: wzheng1@bjtu.du.cn

Abstract: With the rapid development of China train operation and control system, validity and safety of behavioral functions of the system have attracted much attention in the railway domain. In this paper, an automated test sequence optimization method was presented from the system functional requirement specification of the high-speed railway. To overcome the local optimum of traditional ant colony algorithm, the maze algorithm is integrated with the ant colony algorithm to achieve the dynamical learning capacity and improve the adaptation capacity to the complex and changeable environment, and therefore, this algorithm can produce the optimal searching results. Several key railway operation scenarios are selected as the representative functional scenarios and Colored Petri Nets (CPN) is used to model the scenarios. After the CPN model is transformed into the extensible markup language (XML) model, the improved ant colony algorithm is employed to generate the optimal sequences. The shortest searching paths are found and the redundant test sequences are reduced based on the natural law of ants foraging. Finally, the Radio Blocking Center (RBC) test platform is designed and used to validate the optimal sequence. Testing results show that the proposed method is able to optimize the test sequences and improve the test efficiency successfully.
Keywords: Ant colony algorithm, maze algorithm, test sequence, optimization, CPN model

1 Introduction

As the safety-critical system, the Chinese Train Control System Level 3 (CTCS-3) is capable of ensuring the safe and efficient operation for a train with high speed and density. Before the system comes into service, it is essential to execute series of tests, including laboratory testing, field testing, integrated testing, and interoperability test to verify the consistence of the system to the requirement specification [1]. A timely and complete test can significantly contribute to finding the drawbacks in the system design and to assuring the appropriate functional behaviors of the system. Taking the "7.23" accident in China as an example, there existed flaws in the fault-tolerant design of the acquisition drive unit, with the result that the Train Control Center (TCC) misunderstood the train occupancy information. One of the reasons was that the software functionalities of the TCC were not fully tested because there were not efficient test methods and enough test time for the signal system before it was put into operation. As the safety-critical system, the Chinese Train Control System Level 3 (CTCS-3) is capable of ensuring the safe and efficient operation for a train with high speed and density. There are more and more different test methods for the safety-critical train control system [2]- [3]. Behrmann proposed the method of test generation based on UPPAAL timed automata [2]. Hessel described how to generate real-time verification tool for the test case [4]. Lee referred to a method of black box testing

based on the input and output finite state machine (I/O FSM) [5]. Zhao proposed the input and output automata of port labeled (LPTIOA), which automatically realizes the test sequences generation for the LPTIOA model validated by UPPAAL of CTCS-3 on-board system [6]. Jaafar put forward a kind of automatic generation method of testing script based extensible markup language (XML) for the characteristics of safety critical system such as large script in scale, complex structure, difficult maintenance and safety testing demand [7]. Although these methods have achieved some success, it still exist many defects .For instance, the concurrent behavior of the system cannot be described by test generation method based on timed automata. Because the generation of test cases in the railway domain has high levels abstraction, automation can not to be realized. In addition, most current test methods are semiautomatic, and the test cases and test sequences manually generated have the drawbacks of a heavy workload, low efficiency, and high demand for professional expertise.

Ant colony optimization (ACO) is employed to find optimal path in the graph theory. It is proposed by Marco Dorigo in his doctoral thesis in 1992 and the idea comes from the behavior that ants search food to find the path [8]. Experiments show that the algorithm has strong robustness and the ability of finding optimal solution. However, there are still some defects in it such as slow convergence speed and stagnation behavior. Ant colony algorithm has attracted wide attention of scholars and the new ant colony algorithm has been widely applied to various fields, such as fault identification, TSP problem and the construction of roads. At present the study of ant colony algorithm is mainly focused on the improvement and application of algorithm. Wu [9] made full use of the concision and efficiency of "2 - exchange method" by leading in mutation mechanism to the basic ant colony algorithm and proposed an ant colony algorithm with mutation features. Wang [10] put forward the meeting algorithm based on ant colony algorithm, improved ant touring quality, combined meeting algorithm with the parallel segmentation strategy, and proposed a segmentation algorithm for TSP problem based on ant colony algorithm.

Taking all these aspects into consideration, a type of improved ant colony algorithm is proposed in this paper to deal with the problems of inefficiency and repetition of the test sequences in railway domain. The ant colony algorithm and maze algorithm are integrated to generate the executive optimal sequences automatically. They are applied to generate Extensible Markup Language (XML) test cases and XML test sequences respectively. These test sequences are totally feasible for practical test executions. In addition, the repeatability rate of the generated test sequences is reduced. Finally, four scenarios from the CTCS-3 system specification are taken as the example and modeled with Petri nets. The algorithm is validated on the radio block center (RBC) test platform.

The rest of this paper is organized as follows. In Section 2, the basic definitions of Petri nets are introduced and the improved ant colony algorithm is specified. In Section 3, the Colored Petri nets are used to model the typical operation scenarios. After the reachable graph of the Petri nets model is analyzed, the test generation software tool is presented. Finally, the generated test sequences are validated by the RBC test platform. Some concluding remarks are presented in Section 4.

2 Automated test generation

The automated test method by CPN model can be depicted as Fig.1 and it is mainly divided into three stages including modeling, test case generation and test sequence generation. From the system requirement specification and the modeling rules, the CPN model is built for the operation scenarios, and then the XML file and the reachable graph of the model are obtained. On the basis of CPN model with XML format, the proposed improved ant colony algorithm is applied

into generate the set of original test cases. Test subsequence can be generated by concatenating the test cases using ant colony algorithm and the information in test cases. Finally, the set of original test cases and the test sequences can be transformed into files with XML format, and then the XML-format test sequences can be used directly in practical test process.

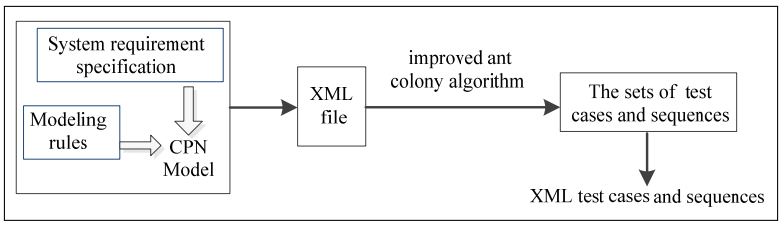


Figure 1: Frame diagram of the test generation method

2.1 Introduction of Colored Petri nets

Colored Petri nets (CPN) is an advanced network system in 1981 introduced by the Jensen Kurt based on basic Petri nets [11]. It is mainly composed of the place, transition, token and arcs to establish a system model. The basic Petri nets has the theoretical basis as well as the wide application in different domains, but it has no data and level concepts so that it is hard to model the complex systems because the model will be very large and difficult to analyzed. CPN extends the basic features of basic Petri nets and introduces the colored markers, arc description, level page and so on, to compensate the drawbacks of the base Petri nets. CPN has the description capacity for the concurrent, synchronous, asynchronous features, resources competition and the coordination of the discrete event dynamic system behavior. If it is combined with the high-level programming language, it can model the large-scale system with the definition of complex data types and data processing.

2.2 Ant colony algorithm

As the ant colony search for the food, they can usually find an optimal path from the nest to food because the ants in the searching path will release a special kind of pheromone. When they encounter a road they never meet before, they randomly select another path. At the same time, the ants release a kind of pheromone associated with the path length. The longer the path, the lower the concentration of hormone released. If the following ants meet this intersection again, the probability that they choose the path with higher hormone concentration will be relatively large. This forms a positive feedback mechanics. The hormone concentration of the optimal path is stronger and stronger, while the other path will diminish as time goes by. Eventually the whole colony will find the optimal path. In addition, the ants can adapt to the environment change. If the obstacle suddenly appears on the moving route of the ants, they can quickly find the optimal path. During the process of searching for the route, although the selecting ability of single ant is limited, the path information is exchanged with the effect of the hormone between different ants, and therefore the ant colony could finally find the optimal path.
There exists other path searching algorithm such as traditional depth-first searching (DFS) algorithm and Bee colony algorithm. Although DFS algorithm is able to search in different levels [12], it doesn?t take recurring nodes into account so that it easily leads to an endless loop, leading to the instability of the searching process. Bee colony algorithm is a type of optimization method by imitating the behavior of bees and the main characteristics of the algorithm is it need not know specific information and only compare the merits of the issue. However, the bee colony optimization algorithm method is too restrictive and it can achieve some success in local optimization,

but from a global point of view, it is not stable enough and has the relatively large blindness. The principle of genetic algorithm, similar to neural network algorithm, uses eugenics survival of the fittest way and achieves the optimization from generation to generation. The neural network mainly finds the optimal way by training, resulting in the relatively low convergence speed.

2.3 The improved ant colony algorithm

The traditional ant colony algorithm is prone to reach the local optimum. In this paper, the maze algorithm and ant colony algorithm are integrated to overcome this shortcoming. For maze algorithm, facing a fork in the maze, the mice can choose one of the roads. If the road is blocked, the mice will return to the fork and walk the other road. In fact, this idea is a process of learning and optimization, which can adapt to the complex and changeable environment. Therefore, together with the idea and the traditional ant colony algorithm, the solutions for the test sequence optimization are designed.

Assuming that the artificial intelligence ant can secrete a special kind of pheromone, if it finds no food when meeting a node, it can inform the foraging ants that there is no food. The ant returns to the nest (initial point) and informs the ant in the nest that this road is blocked and other roads can be selected. If one of the ants can go straight from the initial point to the end point, this road should be recorded as one of the test sequences. In order to prevent the local optimization, the next foraging ants take different roads from those recorded. If other roads can be found, the recorded optimal road is desired. The reverse thinking can be used to accomplish the aforementioned idea quickly. For one complex scenario, the number of roads is limited, so the test sequence sets can be set up. Those test sequences with nodes containing no food should be deleted from the test sequence set. Among those remaining test sequences, the optimal sequence is the one with the minimum nodes. The reverse thinking can greatly reduce the program running time. In this paper, it is assumed that each node has two different states, in which "0" means there is no food here whereas "1" means this node has food.

From the aforementioned principle, the flow diagram of improved ant colony algorithm is shown as Fig.2 [13].

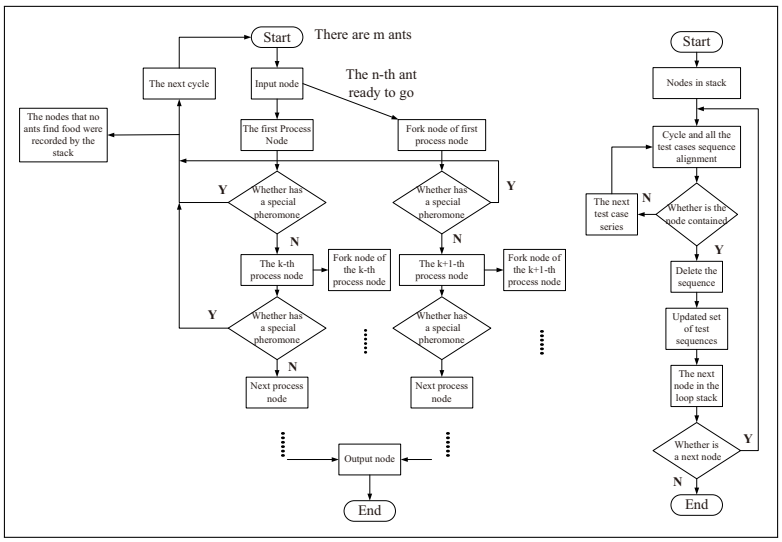


Figure 2: The flow chart of improved ant colony algorithm

3 Application of the test generation algorithm

Radio Blocking Center (RBC), as key trackside equipment in CTCS-3, can generate the movement authority (MA) and other control commands according to the information coming from other subsystems [1]; hence, it should satisfy the functional requirements in the CTCS-3 specifications. Four operation scenarios associated with RBC will be modeled using the CPN and the corresponding test cases and sequences will be generated using the proposed path optimization algorithm. Compared with the RW-TSG [14] and traditional DFS algorithms, the numbers of the test cases and sequences will be seen as the indicators showing the superiority of the proposed algorithms.

3.1 Functionality of RBC

The main functionalities of the RBC include: (1) receiving the position and train data from the onboard equipment through the global system for mobile communications for railways (GSM-R) wireless communication system; (2) generating the MA according to the information from the Balise, the track circuit, the temporary speed restriction server, and the interlocking system; (3) transmitting the MA to the onboard equipment, which will calculate the location-speed curve to protect the train safety operation [15]. In CTCS-3, the RBC is connected with the centralized traffic control (CTC) system, the computer-based interlocking (CBI) system, the temporary speed restriction server (TSRS), and the neighbor RBC (NRBC) by different specific safety data communication "Ethernet," and it is connected with the onboard equipment by GSM-R.

3.2 The main operation scenarios

The main operation scenarios of CTCS-3 includes registration and start, log out, level conversion, movement authority, RBC handover, temporary speed restriction and so on. In this paper, only the registered and start, RBC handover, automatic neutral, and log out are studied.
1. Registration and start: describe the process of the train from the initial location to prepare operation. It mainly includes 6 stages, such as train awaking, train registration, inputting train data and so on.
2. RBC handover: guarantee the train running from one RBC region to another RBC region. The neighboring RBCs should have direct communication to exchange RBC handover information. Schematic diagram is shown as Fig.3.
In the process of RBC handover, RBC1 (handover RBC) is responsible to send the handover warning information (ID, the border of RBC Balise group ID, train data) , route request information, switching notice information, switching confirmation information and so on. RBC2 is responsible to send route information to take over train information to RBC1.

3. Automatic passing over of neutral section: the operation scenario of phase area has no electric interval of the electrified railway. For CTCS-3, RBC sends the phase zone information and movement authorization (MA) to train, and the zone information includes the distance to phase point, the length of phase zone and so on. The schematic process is shown in Fig.4.
On-board equipments receive the activated phase information transmitted by RBC, supervises speed and position of the train in real time, process as followed: 1) When there are 10 seconds running distance from the neutral section to the front of the train, the on-board equipment warns the driver;2) When there are 3 seconds running distance from the neutral section to the front of the train, the on-board equipment sends neutral-section passing order to Electric-Multi-Unit(EMU); 3) After the front of the train crossing the phase region, the related commands are canceled by train equipment.

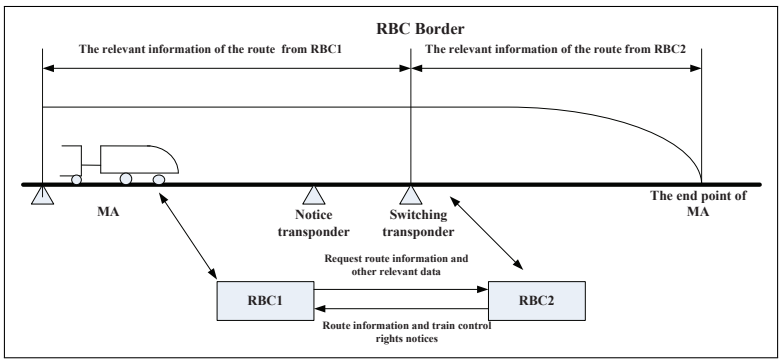


Figure 3: Schematic diagram of RBC switching

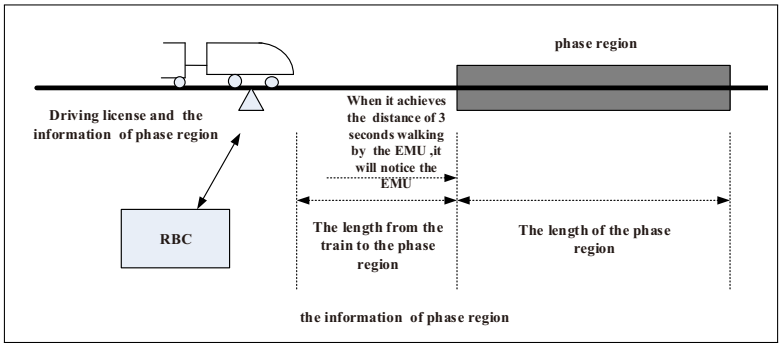


Figure 4: automatic passing over of neutral section

4. Logout: describes the process from logout information to shut off the power for the train. It mainly includes 2 stages: RBC logout and the power of the train is cut off.

3.3 RBC scenarios modeling and test optimization

From the functional characteristics of RBC and the overall technology scheme of CTCS-3, the RBC test cases of XML format and state space reachable graph are established. By the improved ant colony algorithm, several cases are optimized and generated to test sequences with XML format. Finally, the set of the generation test sequence with XML format was validated by RBC simulation testing platform.

1. Operation scenarios modeling

Taking the four scenarios as the example including "Registration and start", "RBC handover", "Automatic passing over of neutral section" ,"Logout", the four operation processes are modeling with CPNs and the high level model are presented by Fig.5 and the other low level mode are described as Fig.6 to Fig.8.

2. Reachable state analysis of the Petri nets model

According to the four functional scenarios selected and the relevant requirement specifications of CTCS-3, the state space reachable graph can be generated with the XML from the CPN models. Some of those redundant states are deleted. After simplification, the simplest 36 states can be

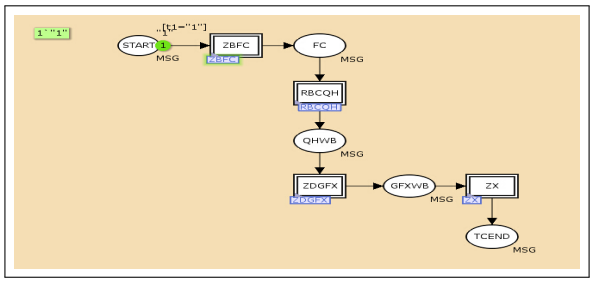


Figure 5: High-level model for the four operation scenarios

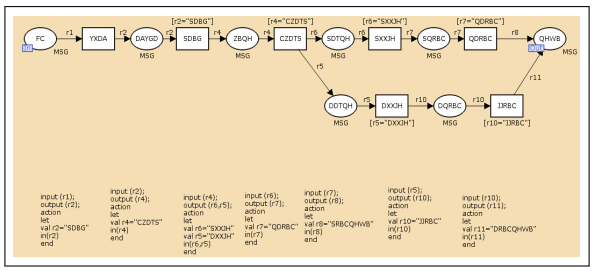


Figure 6: RBC handover scenarios

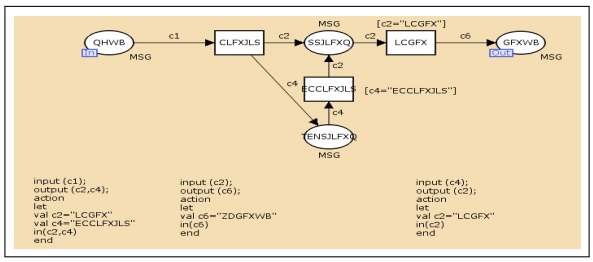


Figure 7: Automatic passing over of neutral section scenario

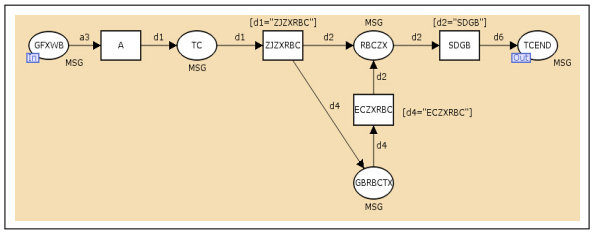


Figure 8: Registration and start scenario

attained through the equivalent markers, as shown in Fig.9.

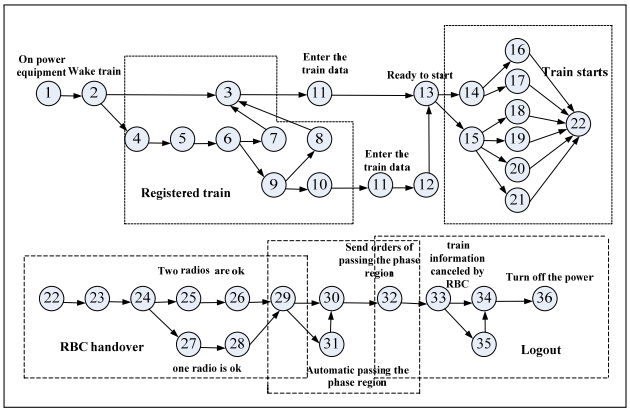


Figure 9: The state diagram of four scenarios

The input, the output nodes of four scenarios are shown in Tab.1.

Table 1: Description of the scenario

Number	Scenario	Input node	Output node
1	Registration and start	1	22
2	RBC switch	22	29
3	Automatic passing the phase region	29	32
4	Logout	32	36

As is shown in Table 1, the input node and output node of scene 1 is 1 and 22 respectively and so on. In order to ensure that all nodes are covered, between two adjacent, input nodes and output nodes have overlapped.

3. The theoretical analysis of the state diagram

From a simplified state diagram of Fig.9, one of the test sequences is "1 2 3 11 13 14 16 22 23 24 25 26 29 30 32 33 34 36". Assuming that the ant A finds "no food" in node 3 during foraging process, which means the next ant need not go into the node. Then ant A goes back to the initial point to exchange the information to ant B. Node 3 is no food and then ant B will go to node 6 and finds that fork in the road. Ant B comes to node 7, finding that the next node is node 3, and then return to the initial point and record that node 7 should not go, until the ant reaches the end. Assuming that in a certain state, node 3, node 15, node 17 and node 25 are dead, according to the above process, we know that the remaining 4 test sequences, such as "2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 30 32 33 34 36", "1 2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 31 30 32 33 34 36", "1 2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 31 30 32 33 35 34 36", "1 2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 30 32 33 35 34 36". From these four test sequences, in order to find the shortest test sequence, after comparing one by one, sequence "1 2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 30 32 33 34 36" will be identified and this test sequence contains the least test cases.

From the simplified state space reachable graph, the program may used to generate the test cases, as shown in Table 2.

From the results, the scenario 1 generates test subsequence ①(1 2 4 5 6 9 10 11 12 13 14 16 22); test subsequence ②(22 23 24 27 28 29) is generated by scenario 2; test subsequence ③(29 31 30 32) and test subsequence ④ (29 30 32) are for the scenario 3; test subsequence ⑤(32 33 34 36) and test subsequence ⑥(32 33 35 34 36) are for the scenario 4. Based on the improved ant colony

Table 2: Test subsequences of the scenarios

Scenarios	Generated test subsequences
1	1 2 4 5 6 9 10 11 12 13 14 16 22
2	22 23 24 27 28 29
3	29 31 30 32/29 30 32
4	32 33 34 36/32 33 35 34 36

algorithm proposed in this paper, the test subsequences combinations by the scenarios are shown in Tab.3:

Table 3: The test subsequences combinations

The number of subsequence	The test subsequences combinations
1	① ② ③ ⑤
2	① ② ③ ⑥
3	① ② ④ ⑤
4	① ② ④ ⑥

After those test sequences is optimized, the final optimization test sequences are generated and shown as Table 4.

Table 4: Optimization of the test sequences

The number of sequence	The test subsequences combination	The test sequences
1	① ② ③ ⑤	1 2 4 5 6 9 10 11 12 13 14 16 22 23 24 27 28 29 30 32 33 34 36

From the aforementioned data, the experimental results conform to the theoretical analysis. By the improved ant colony algorithm, four test sequences are generated and optimized to one eligible case sequence with the minimum redundancy. The test results are real and effective, which not only saves time but also improve the test efficiency greatly.

3.4 Test generation software tool for the ant colony algorithm

This test generation software tool is achieved by C++ language. It executes the ant colony algorithm and generates the test sequences. Fig.10 shows the main interface of the software and it reads the XML model from the CPN model and shows the test sequence generation from several functional scenarios.

If the top right button "reads the XML file" of the main interface is pressed, Fig.11 will appear and show the contents of the XML files to be read.

If natural number greater than 36 is input to the top left four boxes of the screen, Fig.12 presents the optimized test sequences generation.

3.5 Test generation method validation

From the rules of all modules functional requirements of CTCS-3 RBC test that are made by the Ministry of Railways, the RBC test platform is designed and the platform structure is shown in Fig.13.

In order to verify the practicality and feasibility of the test automated generation method, an RBC functionality test platform is developed and the generated test cases and sequences are

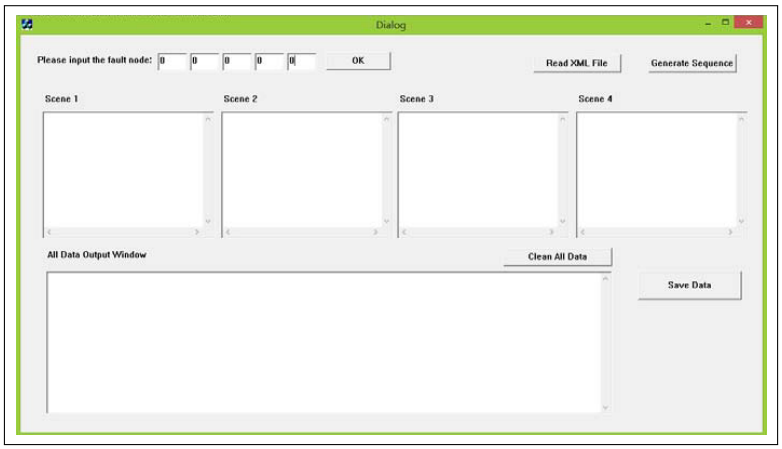


Figure 10: The main interface of the test generation software tool

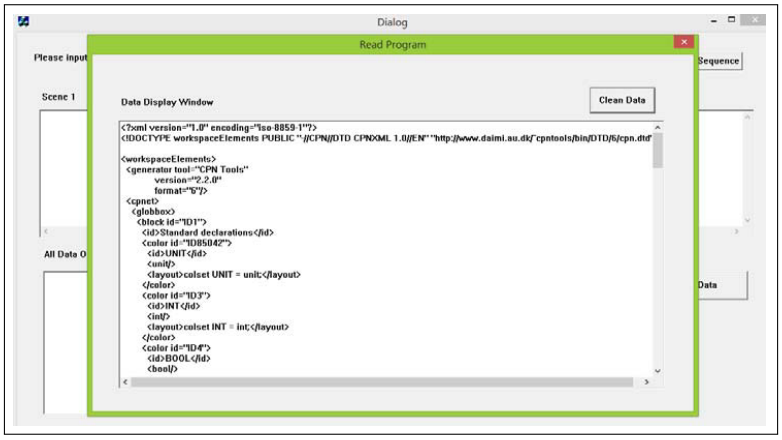


Figure 11: Contents of the XML files

Figure 12: Test sequence optimization generation interface

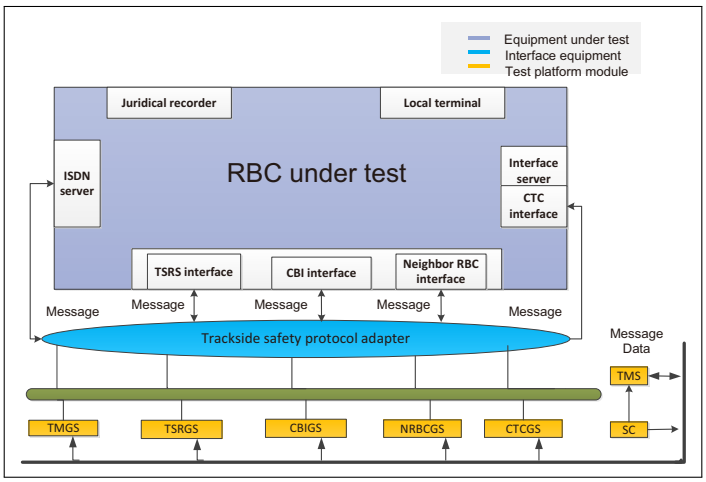


Figure 13: Framework of the RBC test platform

injected to test.

This RBC test platform consists of scenario controller (SC), the centralized traffic control generation simulator (CTCGS), the train message generation simulator (TMGS), the computer-based interlocking generation simulator (CBIGS), the Neighboring RBC (NRBC) generation simulator (NRBCGS), the temporary speed restriction generation simulator (TSRGS), and the train movement simulator (TMS). Five subsystems (except the SC and the TMS) only have message interaction with the RBC under test, and six subsystems (except the SC) have no communication with each other. The SC can control the whole test process and the data for these six subsystems during the test process come from the SC and the SC can directly read the XML test cases and sequences files to get the test data.

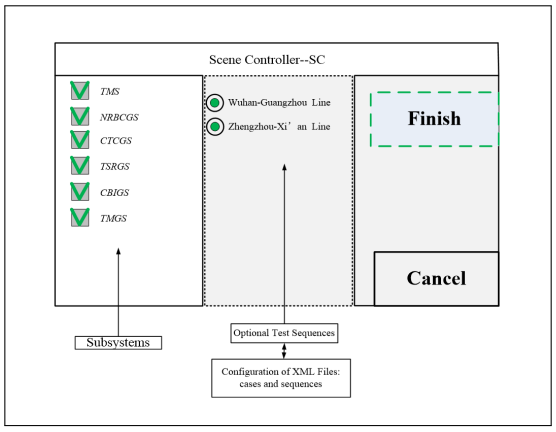


Figure 14: The control interface screen of the SC

The control interface screen of the SC is shown in Fig.14. During the test process, the XML files has the distinguished characteristics, such as understandability, platform independence, the formatted structure, and the ability of hierarchical data description [16] . /citeli and it contains all the test data and can be directly read by the SC, which makes the test process automatically. If a test begins, the involved subsystems and the optional test sequence should be selected first. When the button "Finish" is pressed, the SC will load the XML configuration document and distribute the data to the corresponding subsystems. Then, each subsystem begins to communi-

cate with the RBC under test until the test process is an error or until the test is normally over. The optimal test sequence with XML format generated by the proposed algorithm is applied in the RBC test platform. Part of the key interaction interface data are shown in Tab.5, including the interaction message from the operation scenario "registration and start", "RBC handover" and "Logout"

Table 5: Test data of the RBC platform validation process

Operation scenario: Registration and start							
Data_id	TestSeq_ID	Time	Event	Direction	Msg_ID	Meaning of Msg_ID	MsgData_bit
1	1	2014/11/23/11:48:01	Running	Train→RBC	155	Communication begins	9B0280000019020C01BF
2	1	2014/11/23/11:48:02	Running	Train←RBC	32	System version	2002C00000191FFFFFE4
3	1	2014/11/23/11:48:03	Running	Train→RBC	146	Confirmation	920380000EB4020C0180
4	1	2014/11/23/11:48:04	Running	Train→RBC	159	Communication sets up	9F0280000E90020C01BF
5	1	2014/11/23/11:48:06	Running	Train←RBC	24	General information	1805C0000E903FFFFFE7
6	1	2014/11/23/11:48:07	Running	Train→RBC	129	Train data	8109C0000EB4020C0180
7	1	2014/11/23/11:48:07	Running	Train←RBC	8	Train data confirmation	080380000EB41FFFFFE0
Operation scenario: RBC handover							
Data_id	TestSeq_ID	Time	Event	Direction	Msg_ID	Meaning of Msg_ID	MsgData_bit
1	1	2014/11/23/11:51:50	Running	Train←RBC	3	Movement authority	0327000021F631E0E381F
2	1	2014/11/23/11:51:51	Running	Train→RBC	136	Train location	880600002374820C048001
3	1	2014/11/23/11:54:52	Running	NRBC→RBC	201	Handover announcement	C90020023C00010830060
4	1	2014/11/23/11:54:53	Running	NRBC←RBC	205	Reply	CD0523C000820C01A3C0
5	1	2014/11/23/11:54:54	Running	NRBC→RBC	202	Movement authority request	CA0018023C00010830060
6	1	2014/11/23/11:54:54	Running	NRBC→RBC	205	Reply	CD0523C000820C01A3C0
7	1	2014/11/23/11:54:55	Running	NRBC→RBC	221	Information request	DD2B23C000820C01A3C
Operation scenario: Logout							
Data_id	TestSeq_ID	Time	Event	Direction	Msg_ID	Meaning of Msg_ID	MsgData_bit
1	1	2011/11/23/12:11:45	Running	Train→RBC	156	Communication terminates	9C06000065ED820C02BF
2	1	2011/11/23/12:11:46	Running	Train←RBC	39	Confirmation	2702800065ED91E0621F2
3	1	2011/11/23/12:11:47	Running	Train←RBC	39	Confirmation	2702800065ED91E0621F3

In Table 5, "Msg_ID" records the message ID, and "MsgData_bit" records the content of messages. Taking operation scenario "registration and start" as the example, firstly, the train send Message 155 to RBC, presenting "the communication begins"; Then the RBC send Message 32 to the train, informing the "system version" . After the train confirms the same system version with Message 146, it sends Message 159 to RBC, meaning "the communication sets up". Then the RBC sends general Message 24 to train, presenting the successful communication. Finally, the train set train data Message 129 to RBC and after the RBC uses Message 8 to confirm the train data, the train may start. Thus, the practicality and feasibility of the test automated generation method are verified.

4 Conclusion

This paper has put forward a more efficient improved ant colony algorithm integrated with the maze algorithm, which not only completes the automatic generation of test sequence but also greatly improves the test efficiency. The RBC test platform is designed to validate the effectiveness of the optimal test sequences. Although only four train operation scenarios are employed as the example, the proposed algorithm can be applied on all the other operation scenarios described in the high-speed railway system requirement specification. This approach is able to effectively improve the test efficiency, lower the test difficulty, and achieve the goal of test automation.

Acknowledgment

This paper has been supported by the High-speed Railway United Fund of National Natural Science Foundation of China(U1434209).

Bibliography

[1] System Requirements Specification of the CTCS-3 Train Control System(v1.0). The Ministry of Railways of The People's Republic of China, Beijing: China Railway Publishing House,2008. (in Chinese)

[2] Behrmann, G.; Larsen, K. G. et al (2001); Uppaal-present and future, *Decision and Control, Proceedings of the 40th IEEE Conference on*, 3: 2881-2886.

[3] Samuel, P.; Joseph, A. T. (2008); Test Sequence Generation from UML Sequence Diagrams, *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 879-887.

[4] Hessel, A.; Pettersson, P. (2004); A Test Case Generation Algorithm for Real-Time Systems, *Quality Software, 2004. QSIC 2004. Proceedings. Fourth International Conference on*, 268-273.

[5] Lee, J.D. et al (2007); Verification and Conformance Test Generation of Communication Protocol for Railway Signaling Systems, *Computer Standards and Interfaces*, 29(3): 143-151.

[6] Zhao, X.; Li, K.; Tang, T.; Yuan, L. (2010); Study and Application of UML Based CTCS-3 Operational Scenarios Analysis Approach, *Railway Signalling and Communication*, 8(2): 4-8.

[7] Jaafar, M.F.; Selamat, M.H.; Ghani, A. (2006); TCML-an XML-based test case format, *Computing & Informatics, ICOCI'06, International Conference on*, 1-4.

[8] Dorigo, M. (1992); (1992); Optimization, Learning and Natural Algorithms, Ph.D. Thesis, Dip. Elettronica e Informazione, Politecnico di Milano, Italy.

[9] Wu, H.F.; Chen, X.Q. et al (2013); Improved Ant Colony Algorithm Based on Natural Selection Strategy for Solving tsp Problem, *Journal on Communications*, 34(4): 165-170.

[10] Wang, J.; Wang J. (2008); An Improved Ant Colony Algorithm for Solving tsp Problem, *Computer Technology and Development*, 18(2): 50-52.

[11] Jensen, K. (1994); An Introduction to the Theoretical Aspects of Colored Petri Nets, *Springer Berlin Heidelberg.*

[12] Shaffer, C.A. (2011); Data Structures & Algorithm Analysis in C++, 3rd ed. *New York, NY, USA: Dover*, 390-394.

[13] Hu, N.W.; Zheng, W. (2014); The Improved Ant Colony Algorithm Test Sequence Optimization Based on the RBC Test Platform, *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, IEEE, 2014: 2261-2261

[14] Farooq, U.; Lam, C.P.; Li, H. (2008); Towards Automated Test Sequence Generation, *Software Engineering, 2008. ASWEC 2008. 19th Australian Conference on*, IEEE, 2008: 441-450

[15] *** Functional Requirements Specification of the CTCS-3 Train Control System (v1.0), The Ministry of Railways of The People's Republic of China, China Railway Publishing House, Beijing, China, No. 113, 2008.

[16] Chawathe, S.S. (2004); Real-Time Traffic-Data Analysis, *Intelligent Transportation Systems Proceedings. The 7th International IEEE Conference on*, IEEE, 112-117.

[17] Li, R.M.; Lu, H.P.; Qian, Z.; Shi, Q.X. (2005), Research of in the integrated transportation information platform based on XML, *Intelligent Transportation Systems, 2005. Proceedings*, IEEE, 498-503.

Author index