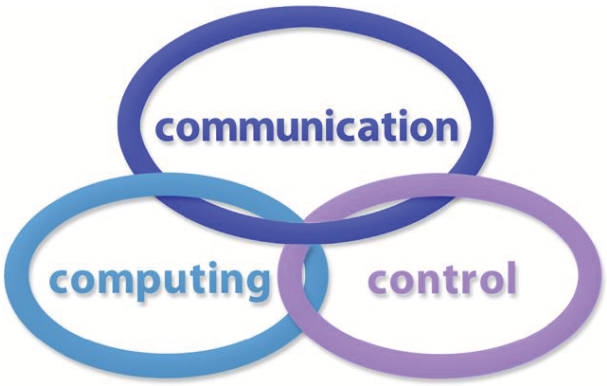
INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836

ISSN-L 1841-9836



A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2014 Volume: 9 Issue: 4 (August)

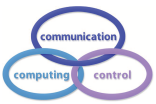
This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

Agora University Editing House

CCC Publications

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

International Journal of Computers Communications & Control



Editorial Address:
Agora University/ R&D Agora Ltd. / S.C. Cercetare Dezvoltare Agora S.R.L.
Piaţa 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/

International Journal of Computers Communications & Control

EDITORIAL BOARD

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
Institute of Intelligent Engineering Systems
Budapest Tech
Budapest, Bécsi út 96/B, H-1034, Hungary
rudas@bmf.hu

Yong Shi
Research Center on Fictitious Economy
& Data Science
Chinese Academy of Sciences
Beijing 100190, China
yshi@gucas.ac.cn
and
College of Information Science & Technology
University of Nebraska at Omaha
Omaha, NE 68182, USA
yshi@unomaha.edu

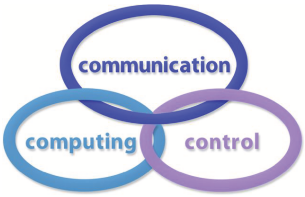
Athanasios D. Styliadis
University of Kavala Institute of Technology
65404 Kavala, Greece
styliadis@teikav.edu.gr

Gheorghe Tecuci
Learning Agents Center
George Mason University, USA
University Drive 4440, Fairfax VA 22030-4444
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
Professor,
Graduate School,
Director,
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division
Department of Electrical Engineering
& Computer Sciences
University of California Berkeley,
Berkeley, CA 94720-1776, USA
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: MILLENNIUM BANK, Bank address: Piata Unirii, str. Primariei, 2, Oradea, Romania
IBAN Account for EURO: RO73MILB0000000000932235
SWIFT CODE (eq.BIC): MILBROBU

International Journal of Computers Communications & Control

Brief Description of IJCCC

Title of journal: International Journal of Computers, Communications & Control
Acronym: IJCCC
Abbreviated Journal Title: INT J COMPUT COMMUN
International Standard Serial Number: ISSN 1841-9836, ISSN-L 1841-9836
Publisher: CCC Publications - Agora University
Starting year of IJCCC: 2006
Founders of IJCCC: Ioan Dzitac, Florin Gheorghe Filip and Mişu-Jan Manolescu
Logo:

Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).

Coverage:

- Beginning with Vol. 1 (2006), Supplementary issue: S, IJCCC is covered by Thomson Reuters - SCI Expanded and is indexed in ISI Web of Science.
- Journal Citation Reports(JCR)/Science Edition:
  - Impact factor (IF): JCR2009, IF=0.373; JCR2010, IF=0.650; JCR2011, IF=0.438; JCR2012, IF=0.441.
- Beginning with Vol. 2 (2007), No.1, IJCCC is covered in EBSCO.
- Beginning with Vol. 3 (2008), No.1, IJCCC, is covered in Scopus.

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 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).

Ştefan Ioan Niţchi

(b. 01-01-1944, Vişeu de Sus, MM - d. 27-04-2014, Cluj - Napoca, CJ)

Editorial Note: We deeply regret the death (April 27, 2014, Cluj-Napoca, Romania) of Dr. Ştefan Ioan Niţchi, professor emeritus and PhD supervisor at Babes-Bolyai University of Cluj-Napoca, distinguished member in Editorial Board of INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL.

Ştefan Ioan Niţchi (born on January 1, 1944, Vişeu de Sus, Maramureş County, Romania) graduated Mathematics (1966) and PhD (1990) in Informatics at Babes-Bolyai University of Cluj-Napoca, Romania. His PhD diploma in Informatics was recognized in USA in 1997. He was 26 years researcher, head of the Programming Environments in Data Bases and Transactional Systems Laboratory at Research Institute for Computer Techniques.

In the time period 1992-1994 was associated professor at Faculty of Economics and Business Administration of Babes-Bolyai University from Cluj-Napoca. From 1994 to 2014 he was full professor. In 1993 was elected Head of Economic Informatics Chair and he was reelected in 1996, 2000, 2004 and 2008. He was in different stages for documentation, technical assistance, senior Fulbright, visiting professor in France, Italy, China, Poland, Czech Republic, USA, Canada, Hungary and Finland.

He published more than 30 books and over 150 papers in journals and proceedings. He has accessed more than 75 research grants. His research activities was focused mainly on Data Bases and Data Warehouses, Decision Support Systems, Artificial Intelligence, Collaborative Systems and E-Business. He was distinguished member of the scientific board of different journals and international conferences.

Mentor and life model, Professor Ştefan Ioan Niţchi lightened and encouraged many young generations with an extraordinary generosity.

Letter from his followers: Professor Ştefan Ioan Niţchi was a man who preserved in his soul the kindness sprung from the awareness of his own value, kindness which was a family inheritance. It was this very kindness that enabled him to approach and look on people and life circumstances with calmness and deep insight. One has to notice that Professor Ştefan Ioan Niţchi has always known to look into the soul of people who he used to work with, which helped him promote valued and talented people and develop long-term relationships based on warm-heartedness in addition to strictly profession-based relationships.

The kindness shown by Professor Ştefan Ioan Niţchi also changed him into a model for many people from inside or outside the academia. He was and will remain a model for the very reason that he always exhibited a conciliatory attitude. He loved to surround himself with harmony. He always attempted to build and maintain a climate of calmness and peacefulness because harmony is what fosters job performance and the much-needed soul balance.

These traits helped him see the good in people around him, praise their qualities and lend an ear to each one in accordance with his own peculiar manner of treating people. Perhaps this is the reason why he gathered so many good people around him who, in turn, learned from their gentle teacher that book learning is not everything in this information-based world of science which is merely an aspect of our life that we boast about and gives us the opportunity to sometimes become conceited and contemptuous. When we remember Professor Ştefan Ioan Niţchi we can only view science as another side of life that

completes the range of other values that allow us to express and act like genuine people among fellow people. It was through kindness that he overcame the desires associated by some people with pride or revenge. Kindness helped Professor Ştefan Ioan Niţchi to maintain the attitude of a modest man, a man who used to talk little and to the point, a man who used to find the best place for everyone.

His behavior gained him the respect of people around him. It is important to highlight that he never demanded or pursued the respect of those around him. He has never been for aggrandizement; instead, people around him always looked up to him as a great professional, a great teacher and a pioneer. Professor Ştefan Ioan Niţchi was the founder of the business IT school in Cluj-Napoca. Throughout his career, he distinguished himself as a professional in the field of databases and decision support systems and proved to be an excellent coordinator of research projects and grants in which he participated. Likewise, the links with businesses as well as with domestic and foreign universities also contributed to the development of the teams that he coordinated.

Last but not least, we find very difficult to put in writing the qualities and deeds that define a man. Indeed, this is a real challenge when it comes to someone with so many accomplishments and who sacrificed his life for the others. If it is true that, when leaving this world, we relive all the feelings we arouse in people around us throughout our entire life, then we are confident that Professor Ştefan Ioan Niţchi relived only warm feelings, the same feelings that he emanated among those who knew him.

We do not claim to have said everything or to have pointed out the most important aspects about Professor Ştefan Ioan Niţchi but we are sure that, each time we reread these lines, he will be dearly remembered.

We will always keep alive in our hearts the image of the MAN and PROFESSOR Ştefan Ioan Niţchi.

Dr. Dan Benţa,
With warmest feelings,
On behalf of Professor's Niţchi disciples/students

International Journal of Computers Communications & Control
ISSN 1841-9836, Volume 9, Issue 4, August, 2014.

Contents

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):389-396, August, 2014.

A Method to Construct Approximate Fuzzy Voronoi Diagram for Fuzzy Numbers of Dimension Two

D. Arotaritei

Dragos Arotaritei
"Grigore T. Popa" University of Medicine and Pharmacy
Romania, 700115 Iasi, Str. Universitatii, 16
E-mail: dragos.arotaritei@umfiasi.ro

Abstract: In this paper, we propose an approximate "fuzzy Voronoi" diagram (FVD)for fuzzy numbers of dimension two (FNDT) by designing an extension of crisp Voronoi diagram for fuzzy numbers. The fuzzy Voronoi sites are defined as fuzzy numbers of dimension two. In this approach, the fuzzy numbers have a convex continuous differentiable shape. The proposed algorithm has two stages: in the first stage we use the Fortune's algorithm in order to construct a "fuzzy Voronoi" diagram for membership values of FNDTs that are equal to 1. In the second stage, we propose a new algorithm based on the Euclidean distance between two fuzzy numbers in order to construct the approximate "fuzzy Voronoi" diagram for values of the membership of FNDTs that are smaller than 1. The experimental results are presented for a particular shape, the fuzzy ellipse numbers.
Keywords: approximated fuzzy Voronoi diagram, fuzzy numbers of dimension two, computational geometry, fuzzy arithmetic, bisector median, path planning.

1 Introduction

Many studies have been made relating to Voronoi diagram and its application [1]. Usually, the main algorithms deal with exact Voronoi diagram. Only few of them deal with approximated Voronoi diagram. In some of paper, 2D objects has been taken into account in order to construct Voronoi diagram. An algorithm to construct Voronoi diagram for convex sites is proposed in [2]. The authors start with the post- office problem where the main question is how to process a set of sites in the plane in order to find the closest site to a query point. The study in [2] treats points as disjoint convex polygons.

In [3] the authors proposed Voronoi diagrams of ellipses using an incremental algorithm that constructs also the exact Delauny graph. A distance from a point to an ellipse is defined to be the minimum Euclidean distance to the boundary of the ellipse with negative value in case the point is inside the ellipse.

Most of the papers treat the problem of disjoint objects. In some papers, the objects can intersect [4] or one object can be inside another object [4]. In the latter case, the circles can intersect each other and circles can contain other circles [4]. In [5] the authors used an incremental algorithm in order to construct Voronoi diagrams for convex objects. A polynomial approximation of the curves is suggested. An algorithm that computes an approximate version of Voronoi diagram is introduced in [6]. The cells in these approximated diagrams are cubes or difference of two cubes. Other papers deal with weighted Voronoi diagrams [7] using the Euclidean distance functions. Each point p from a given set P of points in the plane has an associated weight that expresses the level of influence over this neighborhood.

Fuzzy objects are an important part in the fuzzy set theory. 1-D fuzzy sets have been extended to fuzzy plane and fuzzy plane geometry (circles and polygons) in [9]. Only very few papers deal with fuzzy Voronoi diagram. In ([10]- [11]) the authors proposed an algorithm that deals with fuzzy objects in fuzzy Voronoi diagrams. The "Fuzzy Voronoi" diagram is an extension of the Voronoi diagram based on Fortune's algorithm for crisp cases. The "Fuzzy Voronoi" diagram

has been extended for a set of fuzzy circles, rather than fuzzy points, of the same radius. The authors proved that the boundaries of this diagram are constructed by intersecting a number of hyperbolae and they proposed an algorithm that computes the exact formula [11]. The geometric approach is presented in [12]. The authors define the basic fuzzy geometric shapes (point, line, circle, ellipse and polygon) using the 2-D support space on the level of fuzzy set. A fuzzy measure for trapezoidal fuzzy number is proposed in [13].

Optimal Path Planning for Mobile Robot Navigation is an important area of research in computer science. Usually, obstacles are approximated by rectangular or circular cylinders and Voronoi diagram can be an effective solution [14]. An interesting approach is optimal path planning in an uncertain (fuzzy) environment and 3D space that conduct to idea of fuzzy Voronoi diagram. Fuzzy solution can be very effective for various problems [15].

We present a different algorithm that computes an approximated fuzzy Voronoi diagram for fuzzy numbers of dimension two. It is possible to extend the proposed algorithm to n-dimensional fuzzy numbers using α-cuts.

2 Fuzzy Numbers of Dimension Two

Before we proceed to the main results described in this paper we make a short review of the crisp Voronoi diagram [1] and FNDT [16]. Some proposal for fuzzy distance are investigated by some authors ([17]- [19]). We will use Euclidean distance in fuzzy workspace and some of notations adapted from [11] for Voronoi cell and Voronoi diagram.

Let$P$ be a discrete subset in the metric space X. The Voronoi cell $V(p)$ is defined as the set of all points x in X that have the distance to p lower or equal to the distance from x to other points q in P [11].

$$V(p) = \{x \in X | \forall q \in P, d(x,p) \leq d(x,q)\} \tag{1}$$

The Voronoi diagram of P is the set of all Voronoi cells denoted by $V(p)$. The points p are named Voronoi sites of the set P ([10]- [11]). A fuzzy numbers of dimension n can be defined according to ([9], [16]). Let's consider a fuzzy graph $G \subset R \times R... \times R$, a functional relation in R^n having the membership function $\mu_G(x_1, x_2, ..., x_n) \in R^n$. We define an n-dimensional fuzzy number $F[A]^n$ as any non-empty convex fuzzy subset of $G \in R^n$ that meets the following conditions [16]:

- A is a normal fuzzy set, i.e. there is an $x_0 \in R^n$, $\mu_G(x_0) = 1$;
- A is a convex set, that is membership function is convex hypersurface, i.e. $\mu_A(\alpha x + (1 - \alpha)y) \geq min(\mu_A(x), \mu_A(y))$, $\alpha \in (0, 1]$;
- The membership function is a convex hypersurface that for $\forall a \in R$, there is $\mu_A(x) = a$.

In particular, for n=2 we have fuzzy numbers of dimension two (FNDT), as they are defined in ([9], [16]). A general shape is shown in Fig. 1a. In Fig. b we showed the cut of this surface (top view) for a level $\alpha \in (0, 1]$.

The Fuzzy Voronoi diagram involves computing distances in fuzzy spaces. The results of calculus depend of the metrics used for the fuzzy space.An important point is the selection of the shape of FNDT, derivable in all the points or not: convex polygon, rectangular, square, circle, ellipse or closed curves. In our approach, a continuous convex closed curve of FNDT is considered. The circle or ellipsis is a particular case.

Various metrics in $\tilde{F}(\Re)$ have been proposed in the literature, most of them based on the Hausdorff distance ([18]- [19]). Other authors proposed definitions for measures in the Euclidean

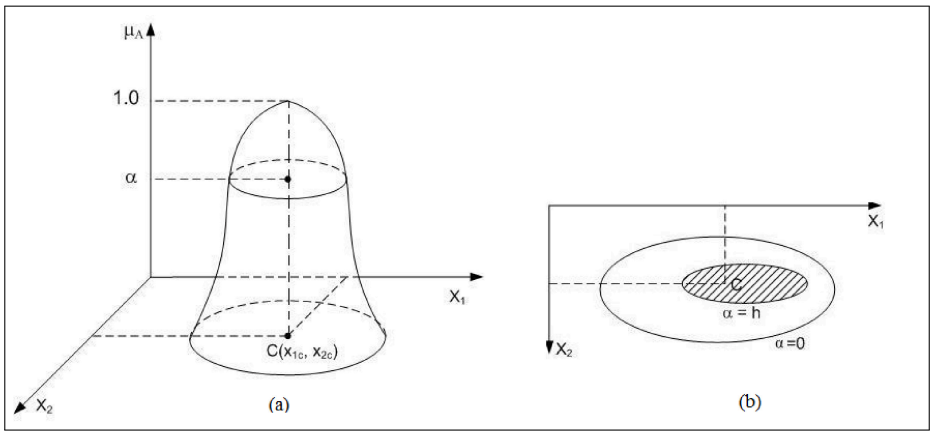


Figure 1: (a) A fuzzy number of dimension 2; (b) Top view of α-cut of FNDT (general case)

space, function that capture more information about vagueness, α-cut that transform fuzzy number into an interval number and distance index based on centroid of fuzzy number. The α-level set A_α of a fuzzy number is a non fuzzy set defined by:

$$A_\alpha = \{x \in \Re | \mu(x) \geq \alpha\}, \quad \mu_A(x) = \sup_{a \in (0,1]} \{\alpha \cdot G_{A_\alpha}(x)\} \tag{2}$$

where $G_{A_\alpha}(x)$ is the characteristic of the function A_α.

3 Approximated Fuzzy Voronoi Diagram (AFVD) for Fuzzy Numbers of Dimension Two

In the next paragraphs we define the FVD for a set of sites that are fuzzy numbers of dimension two. Let $\tilde{p}$ be a subset of $\Re^2$, which is constructed from fuzzy numbers of dimension two and let $\tilde{p} \in \tilde{P}$ be a FNDT. A fuzzy Voronoi cell of $\tilde{p}$ is defined, in a similar manner as in [11], by $\tilde{V}(\tilde{p})$:

$$\tilde{V}(\tilde{p}) = \{(x, \mu_{\tilde{p}}(\tilde{p})) | \forall \tilde{q} \in \tilde{P}, d(x,p) \leq d(x,q)\} \tag{3}$$

We note $p = (p_x, p_y)$ and $q = (q_x, q_y)$ two crisp points that belong to a FNDT at α-cut level. We omitted α from the following formulae in order to simplify the notations. The Euclidean distance between two points p and q that belong to a FNDT at α-cut level is given by:

$$d_E(p,q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \tag{4}$$

This definition is consistent with the results obtained in the particular degenerated case when FNDT become a single point or when $\alpha = 1$. Let us consider two FNDT ellipse shapes not necessary equal. Each level of the α-cuts level will produce two ellipses in the same plane. The fuzzy Voronoi diagram for n FNDT is the union of Voronoi diagrams of fuzzy α-cuts for all n FNDT. Each α-cut of FNDT become a set of ellipses for which a Voronoi diagram is required.

It is easy to demonstrate that given two disjoint continuous convex closed curves A and B, for any two points inside of continuous convex closed curves, the bisector median of these two points are bordered by other two bisector median of two points situated on the contour of the curves. A conclusion of this assertion is that to construct the bisector median of all the points on surface of two continuous convex closed curves it is enough to construct the bisector median for all the points on the contour of the curve that is the border of the bisector median for all the points.

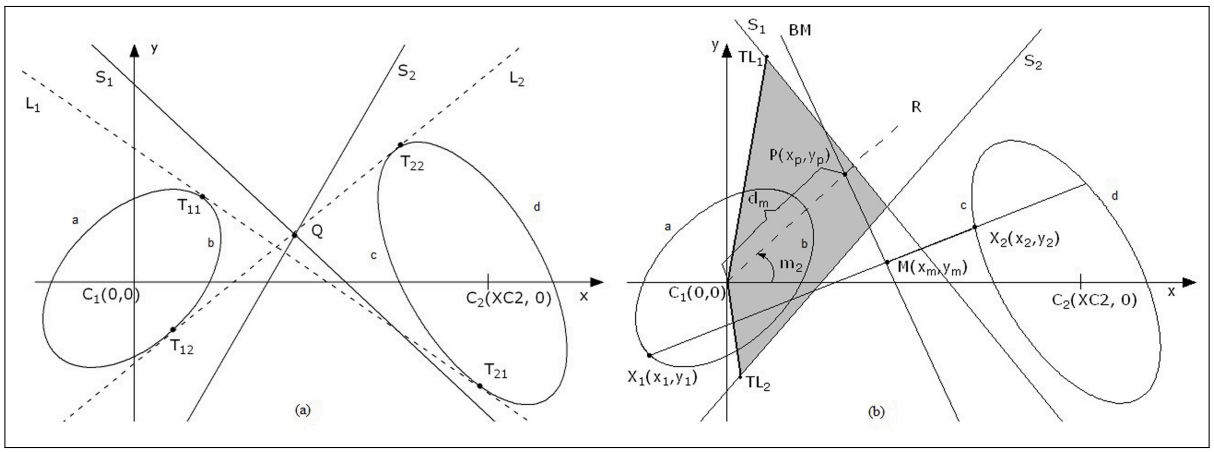


Figure 2: (a) The construction of the bisector median for two points located on the FNDT and the asymptotes of the FVD; (b) The construction of APVD for first curve

For each α-cut, we can compute the fuzzy Voronoi diagram using only the contour of the curve from FNDT obtained by this α-cut. In what follows, we will denote by $h_k \in [0,1)$ the α-cut level of index k.

The starting point for our approach is the calculus of bisector median for two points, one of them situated on the first curve and the second one situated on the second curve. Let's denote by $X_1(x_1, y_1)$ and $X_2(x_2, y_2)$ these two points. The equation of the bisector median is given by a line, perpendicular on the middle of the segment $\overline{X_1X_2}$. If the slope of the line X_1X_2 is m,the slope of the bisector median of $\overline{X_1X_2}$ will be $m' = -1/m = (x_2 - x_1)/(y_2 - y_1)$. The equation bisector median (BM) is a line that pass by the point $M(x_m, y_m)$, the middle of the segment $\overline{X_1X_2}$ and it has the slope m'. After few manipulations, the BMs equation si given by:

$$y = -\frac{x_2 - x_1}{y_2 - y_1} \cdot x + \frac{x_2^2 + y_2^2 - x_1^2 - y_1^2}{2 \cdot (y_2 - y_1)} \tag{5}$$

The fuzzy Voronoi diagram for two FNDT consists from all the bisector medians for all the points situated on the two closed curves. The asymptotes for the envelopes of all bisector medians is given by the bisector median of internal tangents of the two closed convex curves. If we apply counterclockwise rotation and translation for any point with φ angle in Euclidean space, for any two ellipses we have simplified parametric equations of two ellipses (as in Fig. 2(a) and Fig. 2(b)).

Let's consider two convex closed curve that after eventually rotation and translation looks like in Fig. 2(a). The origin of axis is located in the center point of the first unimodal FNDT and his center point of the second unimodal FNDT is located on the axis *0X* (Fig 2(a)).

We consider two points situated on this curves $X_1(x_1, y_1)$ and $X_2(x_2, y_2)$ as in Fig. 2b. The bisector median of these points will be given by formula (5). Let's consider a line that pass in the point $C_1(0, 0)$, with m_2 incrementally from point TL_2 and TL_1 arbitrary chosen on the asymptotic limits for fuzzy Voronoi Diagram S_1 and S_2. The two lines intersect in the point $P(x_p, y_p)$ given by the solution of the system:

$$\begin{cases} y_p = m_1 x_p + n_1 \\ y_p = m_2 x_p + n_2 \end{cases} \tag{6}$$

$$\begin{cases} m_1 = -(x_2 - x_1)/(y_2 - y_1), n_1 = (x_2^2 + y_2^2 - x_1^2 - y_1^2)/(2 \cdot (y_2 - y_1)) \\ m_2 = const., n_2 = 0 \end{cases} \tag{7}$$

$$\begin{cases} x_p = \frac{x_2^2+y_2^2-x_1^2-y_1^2}{2\cdot(x_2-x_1+m_2\cdot(y_2-y_1))} \\ y_p = m_2 x_p \end{cases} \tag{8}$$

If $m_1 \neq m_2$, the two line intersect in the point $P(x_p, y_p)$, where $x_p = (n_1 - n_2)/(m_1 - m_2)$ and $y_p = m_1((n_1 - n_2)/(m_1 - m_2)) + n_1$. In the case $m_1 = m_2$, the two lines are parallel, so no intersection point exists. In our algorithm we simply skip this situation and we go to the next increment.

Let's consider the bisector median for points C_1 and C_2. This crisp Voronoi diagram for two points C_1 and C_2 split conventionally the plane in two halves, left (where first curve is located) and right (where the curve is located). The envelope of the fuzzy Voronoi Diagram for left half-plane is given by the minimum of distance between points P and C_1. For a given m_1, this distance is the intersection of the line R (Fig. 2(b)) and one of the bisector median for all the lines given by formula (5). The minimum of distance is given by:

$$D_m = d^2(C_1, P) = d_m^2 = x_p^2 + y_p^2 = m_1^2 \cdot x_p^2 \tag{9}$$

The function D_m reach the minimum when x_p^2 reach the minimum, that is we must find (x_1, y_1) and (x_2, y_2) for which this minimum is obtained. If the curves are given by parametric equations, this mean that we must find the pair (t_1, t_2) for which $D_m = F(x_1(t_1), y_1(t_1), x_2(t_2), y_2(t_2)) = G(t_1, t_2), t_1 \in \angle(T_{11}, a, T_{12}), t_2 \in \angle(T_{21}, c, T_{22})$ reach the minimum for a given m_2. For each $m_{2j} \in [m_{min}, m_{max}], j = 1, ...n$ we will find an P_j. The union of all the P_j will give the left envelope of the AFVD for these two FNDT at h_k level of α-cut. The envelope is asymptotic limited by the bisector median of the internal tangents.

$$FV_{left} = \bigcup_{j=1}^{n} P_j \tag{10}$$

For the right side, the formulas (6)-(9) are a bit changed, taking into account the equation of the line R that must pass in the point $C_2(XC2, 0)$. After few calculations, the formulas become:

$$\begin{cases} x_p = \frac{x_2^2+y_2^2-x_1^2-y_1^2-q\cdot(y_2-y_1)}{2\cdot(x_2-x_1+m_2\cdot(y_2-y_1))},\ q = 2\cdot m_2 \cdot XC2 = const. \\ y_p = m_2 x_p + n2,\ n2 = -m_2 \cdot XC2 \end{cases} \tag{11}$$

$$D_m = d^2(CA_2, P) = d_m^2 \tag{12}$$

$$d_m^2 = (x_p - XC2)^2 + y_p^2 = (1 + m_2^2) \cdot x_p^2 + 2 \cdot (m_2 n_2 - SC2) \cdot x_p + XC2^2 + n_2^2) \tag{13}$$

AFVD use numerical methods to find the minimum of function given by formulas (7)-(13). Let's suppose that we have the equation of curves given by parametric formula. We can convert the equation of curves into parametric formulas. The quadratic form (circle, ellipse) can be easily converted to parametric form. We taken into account two options. The first one use a minimization with constrains of the function D_m as a function of two parameters in order to find (t_1, t_2) and so one (x_p, y_p).

$$min_{t_1,t_2} F(t_1, t_2) \quad such\ that \quad t_{1,min} \leq t_{1,max}, t_{2,min} \leq t_2 \leq t_{2,max} \tag{14}$$

It is a problem o constrained nonlinear optimization that can be solved by many methods: interior-point, *SQP*, active-set, trust-region-reflective. We chosen to use function *fmincon* implemented in *MATLAB* toolbox.

The second option is possible in the case of twice differentiable curves. We can find the extremum of the function $F(t_1, t_2)$ by calculating the stationary points using first partial derivative and thereafter by select the points that are local minimum using the second derivative. The saddle and the maximum points are eliminated.

Both numerical solutions have a problem: how to choose the initial point in order find the global minimum. The surface of equations (13)-(11) can have many local minima, so we must start with initial solution close enough to global minimum. In order to overcome this situation, we propose to use a brute-force algorithm (BFI) to find the initial points. We split the domain of t_i, i=1,2 in n_i parts. For each combination (t_{ij}, t_{ik}) we find a bisector median and a intersection points. In total we will have maximum $n_1 * n_2$ lines that intersect the line R. Using formula (14), the pairs (t_{10}, t_{20}) for which the minimum of $F(t_1, t_2)$ is obtained is the starting point for minimization in the next refinement stage.

The same reasoning is used for the second curve that have the center $C_2(XC2, 0)$. After both envelopes have been found, we rotate and translate the curves in order to come in initial position. In our experiments we used the first method and the 'trust-region-reflective' algorithm.

4 Algorithm for Approximated Fuzzy Voronoi Diagram and Experimental Results

We can summarize the algorithm for APVD as follow.

Step 1: Construct the crisp Voronoi diagram for crisp points that are the centers of convex closed curves (Fortune's algorithm)P.

Step 2: For each level h_k of α-cuts do the two following steps..

Step 3: For each Voronoi cell that includes a convex closed curve A_i, construct the fuzzy border generated by all adjacent crisp Voronoi cells. The fuzzy border is generated by AFVD described previously.

Step 4: Calculate the intersection of the fuzzy borders and apply the norm max for all fuzzy points.

Step 5: If $h_k \neq 0.0$, go to Step 2, else Stop the algorithm.

In our experimental results we used ellipses given in a parametric form. It is easy to transform the parametric form into quadratic form, and reverse. The experimental results, AFVD for six ellipses, are presented in Fig. 3a and Fig. 3b. In Fig. 3b, the solid line inside the grey area represents the crisp Voronoi diagram for centers C_i of closed curves (ellipses).

Unbounded cells have at infinite the limit determined by internal tangent of adjacent cells corresponding to crisp Voronoi cells of the centers of closed curves (ellipses).

Computational time for Crisp Voronoi (Fortune's algorithm) is given by the known result $O(n \log n)$. Our approach considers a less efficient algorithm, based on construction of bisector lines, for which the computational time is $O(n^2 \log n)$.

5 Conclusions and Future Works

In this paper, we introduced Approximated fuzzy Voronoi diagram for fuzzy numbers of dimension two. We developed an original algorithm for computing fuzzy Voronoi diagram for fuzzy numbers of dimension two with closed convex derivable curve shape. We used α-cuts in order to develop the diagram and the boundaries of Voronoi edges of fuzzy cells.

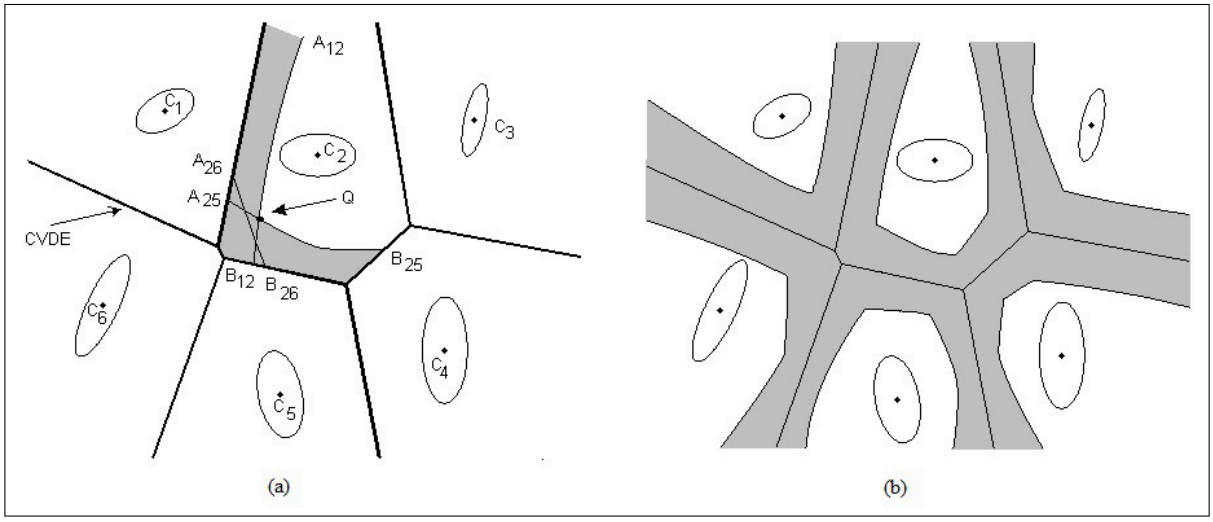


Figure 3: (a)The maximum norm for all fuzzy points inside a fuzzy Voronoi cell (CVDE: Crisp Voronoi Diagram-Edges); (b)AFVD for six ellipses

We considered not only the contour of the fuzzy numbers of dimension two at h level of α-cuts but also points inside this area. The same dimension for all the fuzzy object is a particular case of the presented algorithm. The fuzzy numbers have the same size but different sizes are taken into account.

The experimental results presented the fuzzy Voronoi diagram for FNDT not only for two sites but also for six sites with applicability and algorithm extension to n sites.

Bibliography

[1] Fortune, S. (1887); A sweep algorithm for Voronoi Diagrams, *Algoritmica*, 2:153-174.

[2] Mcallisterm, M.; Kirkpatrick, D.; Snoeyink, J. (1996); A compact piecewise-linear Voronoi diagram for convex sites in the plane, *Discrete Comput. Geom.*, 15-73.

[3] Karavelas, M.; Yvinec, M. (2003); Voronoi diagram of convex objects in the plane, *In Proc. Europ. Symp. Algorithms*, *LNCS Springer*, 337-348.

[4] Yap, C. K. (1987); O(n log n) Algorithm for the Voronoi Diagram of a Set of Simple Curve Segments, *Discrete Comput. Geom.*, 2:365-393.

[5] Arya, S.; Malamatos, T. (2002); Linear-size approximate Voronoi diagrams, *In: 13th Annual ACM-SIAM Symp. on Discrete algorithms*, Society for Industrial and Applied Mathematics, 147-155.

[6] Emiris, I.; Hemmer, M.; Tsigaridas, E.; Tzoumas, G. (2008); Voronoi diagram of ellipses: CGAL-based implementation, *ACS-TR-363603-01 Technical Report*, University of Groningen.

[7] Aurenhammer, F.; Edelsbrunner, H.; An optimal algorithm for constructing the weighted Voronoi diagram in the plane, *Pattern Recognition*, 17(2): 251-257.

[8] Burnikel, C.; Mehlhorn, K.; Schirra, S. (1994); How to Compute Voronoi Diagram of Line Segments: Theoretical and Experimental Results, *In Proc. 2nd Annual Symposium*, Lecture Notes of Computer Science, 855:227-237.

[9] Buckley, J.J. ; Eslami, E. (1997); Fuzzy plane geometry II: Circles and polygons,*Fuzzy Sets and Systems*, 87:79-85.

[10] Jooyandeh, M.; Mohades, A.; Mirzakah, M. (2009); Uncertain Voronoi Diagram, *Information Processing Letters*, 109:709-712.

[11] Jooyandeh, M.; Khorasani, A.M. (2009); Fuzzy Voronoi Diagram, *Advances in Computer Science and Engineering*, 13th International CSI Computer Conference, CSICC 2008, 6:82-89.

[12] Chaudhuri, B. B. (1991); Some shape definitions in fuzzy geometry of space, *Pattern Recognition Letters*, 12: 531-535.

[13] Goetschel, R.; Voxman, W. (1986); Elementary fuzzy calculus, *Fuzzy sets and systems*, 18:31-43.

[14] Takahashi, O.; Schilling, R.J. (1989); Motion Planning in Plane Using Generalized Voronoi Diagrams, *IEEE Transactions on Robotics and Automation*, 5(2):143-150.

[15] Razavi, S.H.; H. Amoozad, H.; Zavadskas, E.K. ; Hashemi, S.S. (2013); A Fuzzy Data Envelopment Analysis Approach based on Parametric Programming, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 8(4):594-607.

[16] Kaufmann, A.; Gupta, M.M. (1984); *Introduction to Fuzzy Arithmetic: Theory and Applications*, an Nostrand Reinhold Company, NY.

[17] Chakraborty, C.; Chakraborty, D. (2006); A theoretical development on a fuzzy distance measure for fuzzy numbers, *Mathematical and Computer Modelling*, 43:254-261.

[18] Grzegorzewski, P. (1998); Metrics and orders in space of fuzzy numbers, *Fuzzy Sets and Systems*, 97:83-94.

[19] Woxman, W. (1998); Some remarks on distance between fuzzy numbers, *Fuzzy Sets and Systems*, 100:353-365.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):397-407, August, 2014.

Petri Net Based Modelling of a Career Syllabus

R. Carvajal-Schiaffino, L. Firinguetti-Limone

Rubén Carvajal-Schiaffino*[1]
Departamento de Matemática y Ciencia de la Computación
Universidad de Santiago de Chile
Av. B. O'Higgins 3363, Santiago de Chile
*Corresponding author: ruben.carvajal@usach.cl

Luis Firinguetti-Limone[2]
Departamento de Estadística
Universidad del Bío-Bío
Av. Collao 1202 - Concepción, Chile
lfiringu@ubiobio.cl

Abstract: A syllabus is a set of courses, their prerequisites and a set of rules defining the continuance of a student in an academic program. The manual verification of the possible contradictions between prerequisites and the set of rules is a difficult task due to the large number of possible cases. In this article we present the different stages in the verification of a syllabus, its modelling and analysis using Petri nets, and suggested ways in which this may be used by the university administration in the decision making process.
Keywords: Decision support systems, system modelling, system verification, educational planning, Petri nets.

1 Introduction

In many countries student retention rates are under scrutiny as part of government funding policy of higher education institutions. As such, retention rates are an increasingly important issue for these institutions ([2]).

It has been found that structural characteristics of higher education institutions, such as enrollment size, selectivity and control, have significant associations with student drop-out (see [5, 15, 16]). In this regard the consistency of a career syllabus may well play a significant role in student persistence/drop-out.

We may define a syllabus as a set of courses, their sequence, the institutional rules defining the continuance of a student in an academic program (number of credits, number of times a course can be repeated, student progress schedule, etc.).

In particular, student progress schedules, which define minimum requirements for continuance in an academic program, are set up by many institutions of higher education (see for instance [13, 14, 17, 18])

Petri net modelling can be an important tool to support decision making at different levels and types of organizations (see for example [4, 6, 10]). In this article we propose a Petri net model of a syllabus, to help the decision making process of the university administrators, students and the course lecturers. The administration may use this model to verify the internal consistency of

[1]Partially funded by DICYT-USACH under grant CS 060933
[2]Partially funded by the Universidad del Bío-Bío under grant DIUBB-083308 3/R

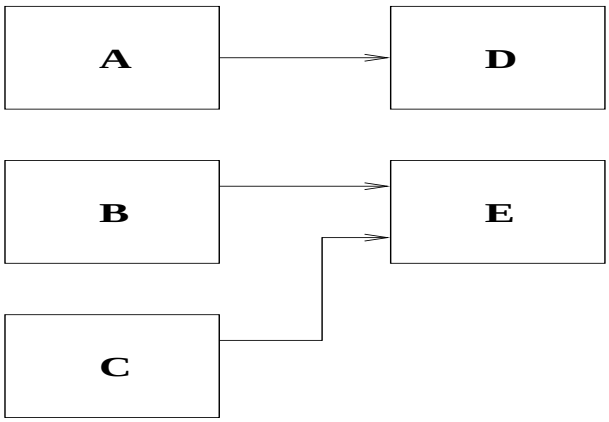


Figure 1: A prerequisites graph

an existing and/or under development syllabus; to plan the student career progress; to establish the risk level of a student drop-out.

In Section 2 we discuss the main characteristics of Petri nets, its feasibility to represent the prerequisites and the analysis of the resulting model. In Section 3 we demonstrate trough an example the stages in the modelling of a rule added to a Petri net representing the prerequisites; an analysis is made of the model obtained and we discuss some related work. Finally, in Section 4 we conclude, presenting our findings and suggestions for future work.

2 Modelling Syllabus Rules

In this article we consider the following rules in a career syllabus:

1.- Prerequisites between courses.

2.- Maximum number of times a course can be repeated.

3.- Maximum duration of stay in the academic program.

4.- Minimum number of courses to be taken during each academic period.

A Petri net [7,8] allows the formal mode description of systems whose dynamics is characterized by *Concurrency, Synchronization, Mutual exclusion, Conflicts*. Prerequisites present all these characteristics, except mutual exclusion. For instance, in Figure 1 we can find concurrency of courses *A, B, C* since, as already mentioned, they have to be taken at the beginning of the program. Another situation of concurrency may occur between courses *B, D* if a student passes courses *A, C* but fails course *B*.

Synchronization is present if to take a course it is necessary to pass more than one, as is the case of *E* since it is necessary to pass *B, C*.

A conflict does not appear explicitly in a prerequisite graph, nevertheless it is present since a course may or may not be passed.

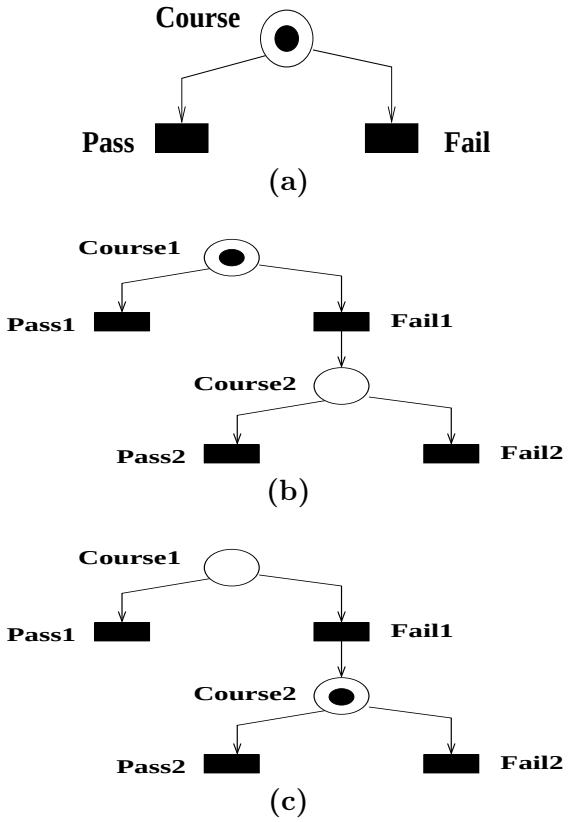


Figure 2: A course

2.1 Modelling Prerequisites and Maximum Number of Failures

Formally, the prerequisites are represented as an acyclic directed graph where each node represents a course and the arcs represent the dependencies (prerequisites) between courses. For example, Figure 1 may represent five courses *A, B, C, D, E* where the first three are the courses to be taken at the beginning of the academic program. The arc joining *A* and *D* pinpoints that to take course *D* it is necessary to pass course *A* and the arcs joining *B, E* and *C, E* indicate that to take course *E* it is necessary to pass courses *B* and *C*.

In Petri nets terms a course is represented by a place, whereas a place with a token indicates that the course is being taken. The activities that can be performed while taking a course are pass or fail. This is represented in Figure 2a.

Since a course can be taken at most twice, this may be represented as in Figure 2b. A token in place Course1 pinpoints that a student is taking a course for the first time. Firing the transition Fail1 indicates that the course was failed, having to be taken again (this is represented by a token in Course2). Firing the transition Fail2 indicates that the student failed Course a second time (Figure 2c), causing its elimination from the program. On the other hand firing transitions Pass1 and Pass2 pinpoint that Course was passed in the first or second instance respectively.

The Petri net representing a complete prerequisites graph is shown in Figure 3. Place *Enrolled* denote the fact that a student is registered in the academic program. Places *A1, A2, B1,*

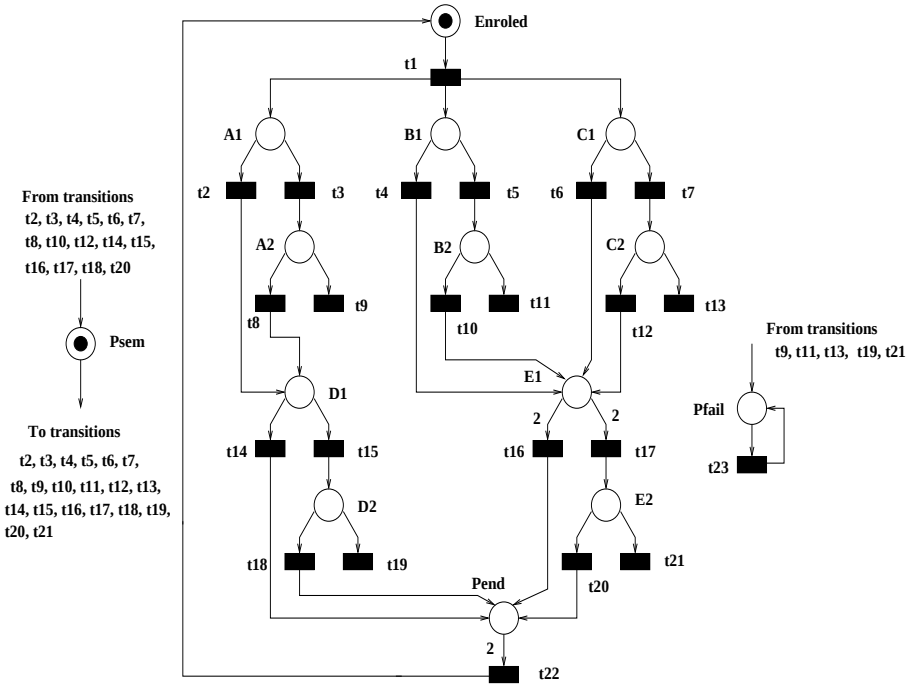


Figure 3: A net representing the prerequisites graph of figure 1

B2, C1, C2, D1, D2, E1, E2 represent different possibilities of passing or failing the courses. The place *Pend* represents a student having taken all the courses of the program. Transition *t1* indicates the beginning of an academic period and firing it puts tokens in places *A1, B1* and *C1*. This global state indicates that a student, once enrolled, must take courses A, B, C. Firing the transitions *t2, t4, t6, t14, t16* respectively represent passing courses A, B, C, D and E the first time these courses are taken. On the other hand, firing the transitions *t8, t10, t12, t18, t20* represents passing the same courses but in the second opportunity. First time failures are produced when transitions *t3, t5, t7, t15, t17* are fired. Second time failures, and elimination from the program, are produced when transitions *t9, t11, t13, t19, t21* are fired.

Since failing a course twice results in the elimination of a student from the program, it is necessary to exclude the possibility of a student taking courses. For this we use place *Psem*, enabling transitions *t2, t3, t4, t5, t6, t7, t8, t9, t10, t11, t12, t13, t14, t15, t16, t17, t18, t19, t20 ,t21*. Transitions *t2, t3, t4, t5, t6, t7, t8, t10, t12, t14, t16, t18, t20* return a token to *Psem*, but firing transitions *t9, t11, t13, t19, t21* represent failing a course a second time, not allowing any further courses to be taken. These failure transitions lead to the state *Pfail* from which it is not possible to escape.

To verify the properties of boundedness and liveness the transition *t22* is fired, putting a token in place *Enroled*.

2.2 Counting the Number of Periods to Conclude a Program

The model discussed on the previous Section allows the direct verification of the conditions of elimination from the program by failing a course a second time. This model satisfies three important properties: boundedness, liveness and absence of deadlock, which are verified by computing P-invariants, T-invariants and generating the reachability graph, respectively.

Algorithm GenGraph
Input: G, a prerequisite graph;
Output: H, a graph of valid courses taken in each period;
1.- Let V, vertex set of H
2.- Add vertex labelled $c_1c_2 \ldots c_k$ to V
3.- Q ← EnQueue(V)
4.- **While** (Q ≠ Empty) **do**
5.- x ← Dequeue(Q);
6.- Let Y be all valid assignments from x taking into account the requisites of G
7.- **For All** $y_i \in Y$ add an arc between x and y_i
8.- EnQueue(y_i)
9.- **EndWhile**

Figure 4: The algorithm.

We have carried the automated analysis of the Petri net model with the tool INA [9].

The number of academic periods necessary to complete the program is taken as a prerequisite for two reasons: to limit the maximum duration of stay in the program and to establish a minimum number of courses to be taken each academic period. Typically, the maximum duration of stay is 1.5 times the duration of the academic program. In our example the duration of the program would be 2 periods and the maximum duration of stay would be 3 periods.

A question to be answered is how to compute the number of academic periods (semesters, years, etc.) required to complete the academic program. To answer this question we must take into account that there are many performance cases. These vary from the optimum case: not failing any course, to the worst case: failing each course once. The model reachability graph contains all these cases, but we must consider that each state represents the firing of a transition not allowing the direct representation of the simultaneous passing/failing of several courses. However relevant information to count the number of periods is given only by states 1, 2, 5 and 7. State 1 (*Enrolled + Psem*) represents a student enrolling in the program the first time; state 2 (*A1 + B1 + C1 + Psem*) represents a student sitting for the first time and simultaneously courses *A, B, C*; state 5 (*D1 + 2E1 + Psem*) represents a student having passed all courses in the first period and sitting now courses *D, E*. The two tokens in place *E1* represents the fact that course *E* has two prerequisites *B, C*. Finally, state *2Pend + Psem* represents a student having taken all program courses.

From the previous discussion it may be deduced that, although the reachability graph contains all the states of the model and gives all valid sequences to complete the program (without time limitations) an additional time effort is necessary to recover only the states representing an academic period. In our example the reachability graph contains 96 states, but only 17 of the are relevant for our analysis.

A Feasible Approach to Counting Academic Periods

This problem can be solved by computing the model t-semiflows because, conceptually, they represent the sequence of fired transitions (in our case passing or failing sequences) to comeback to the initial state. If we regard as the initial state the enrollment of a student in the program and the possibility of taking courses, then any t-invariant will be a sequence of passing/failing

T-Semiflow	Number of Failures	Failed Courses	Number of Periods
t1 + t2 + t4 + t6 + t14 + t16 + t22	0		2
t1 + t3 + t4 + t6 + t8 + t14 + t16 + t22	1	A	3
t1 + t2 + t5 + t6 + t10 + t14 + t16 + t22	1	B	3
t1 + t2 + t4 + t7 + t12 + t14 + t16 + t22	1	C	3
t1 + t2 + t4 + t6 + t15 + t16 + t18 + t22	1	D	3
t1 + t2 + t4 + t6 + t14 + t17 + t20 + t22	1	E	3
t1 + t2 + t5 + t7 + t10 + t12 + t14 + t16 + t22	2	B,C	3
t1 + t2 + t5 + t6 + t10 + t15 + t16 + t18 + t22	2	B,D	3
t1 + t2 + t5 + t6 + t10 + t14 + t17 + t20 + t22	2	B,E	4
t1 + t2 + t4 + t7 + t12 + t15 + t16 + t18 + t22	2	C,D	3
t1 + t2 + t4 + t7 + t12 + t14 + t17 + t20 + t22	2	C,E	4
t1 + t2 + t4 + t6 + t15 + t17 + t18 + t20 + t22	2	D,E	3
t1 + t3 + t5 + t6 + t8 + t10 + t14 + t16 + t22	2	A,B	3
t1 + t3 + t4 + t7 + t8 + t12 + t14 + t16 + t22	2	A,C	3
t1 + t3 + t4 + t6 + t8 + t15 + t16 + t18 + t22	2	A,D	4
t1 + t3 + t4 + t6 + t8 + t14 + t17 + t20 + t22	2	A,E	3
t1 + t2 + t5 + t7 + t10 + t12 + t15 + t16 + t18 + t22	3	B,C,D	3
t1 + t2 + t5 + t7 + t10 + t12 + t14 + t17 + t20 + t22	3	B,C,E	4
t1 + t2 + t5 + t6 + t10 + t15 + t17 + t18 + t20 + t22	3	B,D,E	4
t1 + t2 + t4 + t7 + t12 + t15 + t17 + t18 + t20 + t22	3	C,D,E	4
t1 + t3 + t5 + t7 + t8 + t10 + t12 + t14 + t16 + t22	3	A,B,C	3
t1 + t3 + t5 + t6 + t8 + t10 + t15 + t16 + t18 + t22	3	A,B,D	4
t1 + t3 + t5 + t6 + t8 + t10 + t14 + t17 + t20 + t22	3	A,B,E	4
t1 + t3 + t4 + t7 + t8 + t12 + t15 + t16 + t18 + t22	3	A,C,D	4
t1 + t3 + t4 + t7 + t8 + t12 + t14 + t17 + t20 + t22	3	A,C,E	4
t1 + t3 + t4 + t6 + t8 + t15 + t17 + t18 + t20 + t22	3	A,D,E	4
t1 + t2 + t5 + t7 + t10 + t12 + t15 + t17 + t18 + t20 + t22	4	B,C,D,E	4
t1 + t3 + t5 + t7 + t8 + t10 + t12 + t15 + t16 + t18 + t22	4	A,B,C,D	4
t1 + t3 + t5 + t7 + t8 + t10 + t12 + t14 + t17 + t20 + t22	4	A,B,C,E	4
t1 + t3 + t5 + t6 + t8 + t10 + t15 + t17 + t18 + t20 + t22	4	A,B,D,E	4
t1 + t3 + t4 + t7 + t8 + t12 + t15 + t17 + t18 + t20 + t22	4	A,C,D,E	4
t1 + t3 + t5 + t7 + t8 + t10 + t12 + t15 + t17 + t18 + t20 + t22	5	A,B,C,D,E	4

Table 1: T-semiflows obtained from the net in figure 3

a course such that the last transition fired will be *t22* which allows to comeback to the initial state.

The net possess 33 t-semiflows (Table 1), but only 32 of them are valid for computing the number of academic periods. For instance, the firing sequence necessary to take all the courses in the program failing none is *t1 + t2 + t4 + t6 + t14 + t16 + t22*. With this information it is possible to determine the periods of time taken. For this it is necessary to take into account the places acting as synchronization points, in other words courses with more than one prerequisite. As shown in Table 1, from the total of 32 valid sequences to sit courses without failing more than twice, only 15 allow completion of the program without entering in conflict with the maximum time of continuance in the program. Note that there are situations where failing two courses (*B,E* or *C,E* or *A,D*) results in the elimination of the program.

However, in the case of programs of study with many courses, counting the t-semiflows of the corresponding Petri nets is not possible. The reason for this is that the number of t-semiflows is exponential with respect to the number of courses. For example, for a program of study with 16 courses (4 academic periods), the corresponding reachability graph contains 47,935 states and 65,536 valid sequences for sitting the courses failing at most once some of them.

To solve the problem of the maximum number of semiflows that can be computed with a tool to analyze Petri nets, we propose an algorithm (see Fig. 4) that allows the generation of valid sequences to fulfill a program with no more than two failures in each course. This algorithm generates a directed graph in which each vertex contains a set of courses that can be taken in each academic period.The initial vertex of this graph contains the set of courses to be taken at the beginning of the program. Then, -taking into account the number of resits and the restrictions asserted by the prerequisite graph-, from each vertex are generated the vertices containing the set of courses to be taken next.

Figure 5 shows the graph of the syllabus of 5 courses given in Figure 1; In this case the first vertex, $a_1b_1c_1$, represents courses *A,B,C* which are taken for the first time. The adjacent vertices

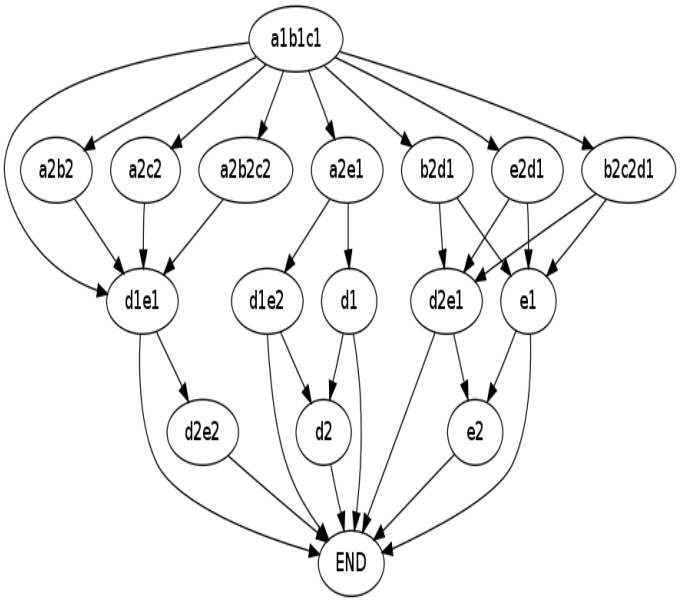


Figure 5: Graph generated with algorithm of figure 4 using the graph of figure 1 as input

represent the courses that are taken in the following period having into account the number of repetitions and restrictions of the prerequisite graph. As can be observed, in this case the number of vertices is 17. A valid sequence of courses taken (having into account repetitions) is obtained if one follows the graph from the initial vertex to vertex *END*. For instance the sequence $a_1b_1c_1, d_1e_1, END$ represents passing all courses, failing none. In the opposite case, the sequence $a_1b_1c_1, a_2b_2c_2, d_1e_1, d_2e_2, END$ represents taking all courses twice. The number of periods taken is given by the number of vertices visited minus 1.

The number of trajectories of all paths in the graph coincides with the number of t-semiflows of the Petri net representing the prerequisites graph with the advantage of greater velocity and permitting the analysis of programs with a larger number of courses.

2.3 Minimum Number of Courses to Be taken During Each Academic Period

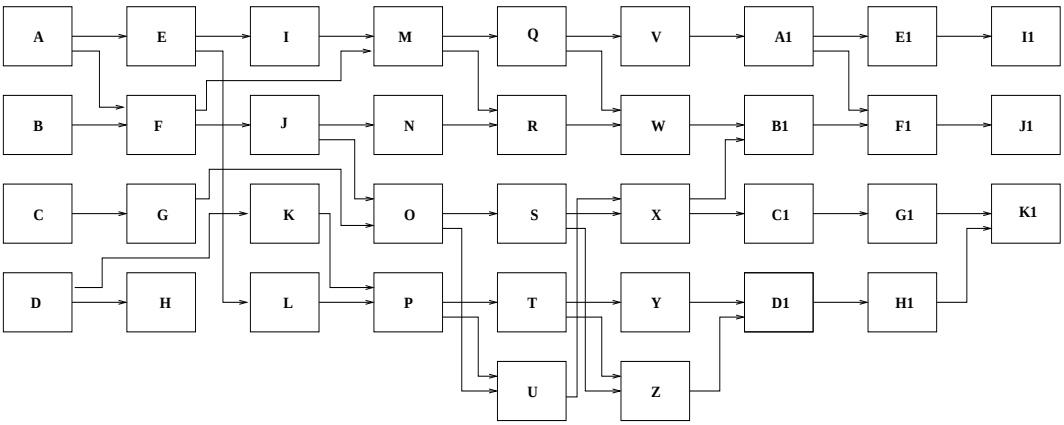


Figure 6: A prerequisites graph for a nine periods program

Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
	2	3	3	3	1

Table 2: Minimum number of courses passed for the academic program

Number of courses taken	Frequency
1	8
2	30
3	72
4	107
5	79
6	24

Table 3: Frequency of the number of courses taken for the academic program

To illustrate the analysis of this rule we use the algorithm of Figure 4, with the first four academic periods of the prerequisite graph of Figure 6. The reduced graph of this model generates 320 states and 65,536 valid sequences. The minimum number of courses passed per semester is given in Table 2.

The number of valid courses taken each semester varies from 1 to 6. Table 3 shows the distribution of the 320 valid sequences. Except for the first, second and sixth period, the minimum number of courses passed per semester is 3. From this, two situations must be analyzed:

a) less than 3 courses are taken

b) more than 4 courses are taken

If less than 3 courses are taken and these are from the last part of the syllabus, no problem will arise, since the student will be taking the courses necessary to complete the program. This will not be the case if the courses are taken before the final part of the program.

For a course taken there are 8 possible cases, discarding the 4 cases that represent taking only courses *M, N, O, P* a second time, there remain four cases to analyze, corresponding to taking courses *F, I, J* a second time or else *P* for the first time.

Path
$A_1B_1C_1D_1 \rightarrow B_2C_2D_2F_1 \rightarrow F_1G_1H_1K_1L_1 \rightarrow F_2$
$A_1B_1C_1D_1 \rightarrow E_1F_1G_1H_1K_1 \rightarrow E_2G_2J_1 \rightarrow I_1L_1N_1O_1 \rightarrow I_2$
$A_1B_1C_1D_1 \rightarrow B_2C_2D_2E_1 \rightarrow F_1G_1H_1I_1K_1L_1 \rightarrow G_2J_1M_1 \rightarrow J_2$
$A_1B_1C_1D_1 \rightarrow C_2D_2E_1F_1 \rightarrow E_2G_1H_1J_1K_1 \rightarrow H_2I_1K_2L_1N_1O_1 \rightarrow L_2M_1N_2 \rightarrow P_1$

Table 4: Sequences of courses taken enabling to take only one course

When it is feasible to take more than 4 courses in each academic period, there are 79 situations with 5 courses (Table 5) and 24 with 6 courses (Table 6). The analysis of these cases are useful to limit the number of courses a student can take, since there is a larger number of situations where a student takes courses for a second time. The proposed method allows the analysis of the student progress schedule (that is to say the total number of courses passed up to a certain period) with the valid sequences generated. This analysis allows to detect if failing a course, which is prerequisite of others, produces the elimination of a student because it may not be possible to take anymore courses to comply with the student progress schedule.

Courses taken for the first time	Total
5	3
4	12
3	19
2	21
1	14
0	10

Table 5: Distribution of courses taken for the first time in groups of five

Courses Taken for the first time	Total
6	2
5	3
4	4
3	5
2	5
1	3
0	2

Table 6: Distribution of courses taken for the first time in groups of six

2.4 Decision Support

Of common interest to university administrators, students and course lecturers is to reduce student drop-out. Our model may help to plan the courses taken by students to achieve this objective.

Valid sequences provided by the previously developed algorithm would help academic administrators, students and lecturers in making decisions. In particular, administrators may use this information to detect which students are at risk of dropping-out and suggest which courses are the most convenient to achieve the objective of completing the academic program.

On the other hand lecturers, once detected students in risks of drop-out, may advice students and administrators to take some corrective actions to reduce this risk.

2.5 Practical Evaluation

We now evaluate the algorithm used to generate the valid sequences of the possible courses to take with no more than two repetitions. Using as entry the graph of prerequisites of Figure 6 we have run the algorithm incrementing the number of levels from 4 to 9.

The results obtained are shown in Figure 7. It can be observed that for the complete syllabus there are 3,838 valid sequences with 93,244,504,769 possible paths. Processing time was almost an hour. The experiments were made with a computer with a 2.8 GHz INTEL processor and 4 Gb RAM.

Another experiment made consisted in counting the number of sequences leading to a situation were the number of courses to take is insufficient, (in our case less than 3). We found 70,207,548,816 cases. We also counted the number of situations were the maximum number of semesters was exceeded. We found 18,475,875,736 of these cases.

Levels	Vertices	Paths	Time
4	464	65,537	0.002s
5	1,184	2,097,153	0.126s
6	2,182	67,108,865	2.497s
7	2,862	882,837,761	29.049s
8	3,440	12,598,171,649	433.505s
9	3,838	93,244,504,769	3,463.607s

Figure 7: Experimental evaluation

3 Results and Discussion

The work presented in [3] uses Petri nets to assist curricula development, modelling the sequence of courses to be taken by students in an academic program. [11] describes the application of model checking to the automatic planning and synthesis of student careers under a set of requirements and to the automatic verification of the coherence of syllabi subject to a set of rules. For this it is suggested to provide the students with a sort of electronic advisor of studies. However, it is recognized that the verification of the coherence of syllabi with the set of rules is by far the most difficult problem and out of the reach of standard technology of information systems, because it requires a (double) exhaustive search engine. The basic step is to explore all possible course combinations to find a feasible set of courses which satisfies the given set of rules and prerequisites to qualify for the university degree chosen by the student. The authors propose to encode and solve the problems of the of the synthesis of student careers and of the verification of the syllabi with respect to the set of rules as computational tree logic model checking problems.

4 Conclusions

We have presented a methodology for the analysis and verification of prerequisites of courses of a syllabus modelled with a Petri net. Also, we have verified that the model is consistent and capable of being generated in automatic mode.

Besides, we have indicated how this model may be used by administrators and lectures to suggest students corrective actions to reduce drop-out caused by a bad selection of course sequences.

The next stage in this research is to enrich the model with the representation of the number credits given to each course. Also an advance table will be included which consists in verifying whether student complies with the minimum number of credits required since he or she enrolled in the program. In a subsequent stage we plan to apply the model to a group of students and produce a quantitative analysis using the stochastic extensions of Petri nets [1].

Bibliography

[1] Ajmone Marsan M., Balbo G., Conte G., Donatelli S., and Franceschinis G. (1995); Modelling with Generalized Stochastic Petri Nets, *Series in Parallel Computing*, John Wiley & Sons.

[2] Doherty W (2006); An Analysis of Multiple Factors Affecting Retention in Web-Based Community College Courses, *The Internet and Higher Education*, 9:245-255.

[3] Ferraris M., Midoro V. and Olimpo G. (1984); Petri Nets as a Modelling Tool in the Development of CAL Software, *Computers and Education*, 8:41-49.

[4] Jahangirian M., Eldabi T., Naseer A., Stergioulas L., Young T. (2010); Simulation in Manufacturing and Business: A Review, *European Journal of Operational Research*, 203:1-13.

[5] Kim D. B.; The Effect of Loans on Students Degree Attainment: Differences by Student and Institutional Characteristics, *Harvard Educational Review*, 77(1):64-100.

[6] Rajabi B. A. and Lee S. P. (2009); Change Management in Business Process Modeling Based on Object Oriented Petri Net, *International Journal of Human and Social Sciences*, 4:13.

[7] Reisig W (1986); Petri Nets. An introduction, *EATCS Monographs on Theoretical Computer Science*, Springer.

[8] Reisig W. and Rozenberg G., editors (1998); Lectures on Petri Nets I: Basic Models, *Advances in Petri Nets*, LNCS 1491. Springer.

[9] Roch S. and Starke P. H. (1999); INA: Integrate Net Analizer, *Humboldt-Universität zu Berlin.*

[10] Salimifard K., Wright M. (2001); Petri Net-Based Modelling of Workflow Systems: An Overview, *European Journal of Operational Research*, 134(3):664-676.

[11] Sebastiani R., Tomasi A. and Giunchiglia F. (2001); Model Checking Syllabi and Student Careers, *Lecture Notes in Computer Science 2031.* Springer.

[12] Silva M., Teruel E., and Colom J. M (1998); Linear Algebraic and Linear Programming Techniques for Analysis of Place/Transition Net Systems. *In Reisig and Rozenberg [8]*, 308-309.

[13] Stevenson University (May 2011); Policies for Continuance and Progression in the Major, www.stevenson.edu/academics/nursing/policies.asp

[14] The College of William and Mary (May 2011); Continuance Requirements for Fulltime Students. www.wm.edu/offices/deanofstudents/policies/academic/contfulltime/index.php

[15] Titus M. A. (2004); An Examination of the Influence of Institutional Context on Student Persistence at 4-year Colleges and Universities: A Multilevel Approach, *Research in Higher Education*, 45(7):673-699.

[16] Titus M. A (2006); Understanding College Degree Completion of Students with Low Socioeconomic Status: The Influence of the Institutional Financial Context, *Research in Higher Education*, 47(4):371-398.

[17] Universitat Pompeu Fabra (May 2011); Regulation Governing the Progression System and the Continuance Rules for Undergraduate Courses, www.upf.edu/universitat/en/normativa/upf/normativa/grau/RD1393/permanencia/

[18] Wabash College. Continuance in College (May 2011); www.wabash.edu/academics/bulletin.cfm?site_code_id=967.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):408-418, August, 2014.

Simulation Experiments for Improving the Consistency Ratio of Reciprocal Matrices

D. Ergu, G. Kou, Y. Peng, X. Yang

Daji Ergu
Southwest University for Nationalities
4th Section,Yihuan Nanlu,Chengdu, 610041, China, ergudaji@163.com

Gang Kou *
School of Business Administration
Southwestern University of Finance and Economics
No.555, Liutai Ave, Wenjiang Zone,Chengdu, 610054, China
*Corresponding author kougang@swufe.edu.cn

Yi Peng
School of Management and Economics
University of Electronic Science and Technology of China
No.2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, 611731, China
pengyicd@gmail.com

Xinfeng Yang
School of Statistics, Southwestern University of Finance and Economics
No.555, Liutai Ave, Wenjiang Zone, Chengdu, 610054, China, 315425159@qq.com

Abstract: The consistency issue is one of the hot research topics in the analytic hierarchy process (AHP) and analytic network process (ANP). To identify the most inconsistent elements for improving the consistency ratio of a reciprocal pairwise comparison matrix (PCM), a bias matrix can be induced to efficiently identify the most inconsistent elements, which is only based on the original PCM. The goal of this paper is to conduct simulation experiments by randomly generating millions numbers of reciprocal matrices with different orders in order to validate the effectiveness of the induced bias matrix model. The experimental results show that the consistency ratios of most of the random inconsistent matrices can be improved by the induced bias matrix model, few random inconsistent matrices with high orders failed the consistency adjustment.

Keywords: Reciprocal random matrix, Consistency ratio, induced bias matrix, simulation experiment; analytic hierarchy process (AHP)/analytic network process (ANP)

1 Introduction

In the analytic hierarchy process (AHP) and analytic network process (ANP), the pairwise comparison matrix (PCM hereinafter) originated by Thurstone [1] is one of the most important components, which is used to compare two criteria or alternatives with respect to a given criterion , then a matrix $A = (a_{ij})_{n\times n}$ is built to reflect the direct and indirect judgment relations between pairs of criteria or alternatives with respect to a given criteria, where $a_{ij} > 0, a_{ij} = \frac{1}{a_{ji}}$. All PCMs are then used to derive the priority vectors, and the alternatives can be ranked by aggregating the local priority vectors [2–4]. However, the decision made based on the final priority vectors is effective only if the paired comparison matrices pass the consistency test [5]. In practice, it is usually difficult to obtain a matrix that satisfies the perfect consistency condition (i.e. $a_{ij} = a_{ik}a_{kj}$ for $i, j, k = 1, 2, \ldots, n$). Therefore, Saaty [6,7] proved that the maximum eigenvalue λ_{max} of matrix A always satisfies $\lambda_{max} \geq n$ and the equality holds if and only if A is perfectly

consistent. Based on this property, Saaty proposed the consistency ratio (CR) to measure the consistency of a matrix, i.e. the consistency of a matrix is acceptable if the CR is less than 0.1. However, this condition sometimes cannot be satisfied with because of the limitations of experiences and expertise, prejudice as well as the complexity nature of the decision problem [8].

To improve the consistency ratio of a matrix, many models and methods have been proposed over the past few decades. For instance, Harker [9] regarded the largest absolute value(s) in matrix $\left\{v_i\omega_j - a_{ji}^2 v_j\omega_i\right\}$ for all i, where $i > j$, as the most inconsistent element(s). Saaty [6] constructed the deviation differences matrix $B = [b_{ij}] = [|a_{ij} - \omega_i/\omega_j|]$ to identify the most inconsistent entry, where ω_i and ω_j are any two subjective priority weights in the $\omega = (\omega_1, \cdots, \omega_n)$. Based on these models, Xu and Wei [10] generated a near consistent matrix $B = (a_{ij}^{\lambda}(\omega_i/\omega_j)^{1-\lambda})_{n\times n}$ to improve the consistency, where λ is a parameter of the auto-adaptive algorithm. Besides, Saaty [7] and Cao et al. [11] introduced Hadamard operator " $\circ$ " to build a perturbation matrix E and a deviation matrix A, in which $E = (a_{ij}) \circ (\omega_j/\omega_i)]$ and $A = [\omega_i/\omega_j] \circ [a_{ij}/(\omega_i/\omega_j)]$, to identify the most inconsistent entry.

There is a common feature in the previously reviewed models, that is, these models are dependent on the priority weights ω_i and ω_j, but there exist more than 20 priority derivation methods [12–14], and the final priority weights obtained from different methods might be different when the matrix is inconsistent. Therefore, Ergu et al. [8] proposed an induced bias matrix to identify the most inconsistent entry in the original inconsistent matrix A. To do so, three major steps containing seven specific steps were proposed and several numerical examples were used to validate the proposed model. In this paper, we attempt to conduct simulation experiments to further validate the effectiveness of the proposed induced bias matrix (IBM) model by generating randomly millions number of the reciprocal positive matrices with different orders. The step 6 and step 7 proposed in Ergu et al. [8] are further quantified and detailed in order to implement automatically modification.

The remaining parts of this paper are organized as follows. The next section briefly describes the induced bias matrix model. The simulation experiments and some algorithms are performed and proposed in Section 3. Section 4 concludes the paper as well as future research directions.

2 The induced bias matrix model

In Ergu et al. [8], the theorem of induced bias matrix and two corollaries were proposed to identify the most inconsistent entries in a PCM and improve the consistency ratio. For the readers' convenience, we first briefly describe the related theorem and corollaries of the IBM model as preliminary of IBM model.

The Theorem 1 says that *"the induced matrix C=AA-nA should be a zero matrix if comparison matrix is perfectly consistent"*. Based on this theorem, if comparison matrix A is approximately consistent, Corollary 1 derived that *"the induced matrix C=AA-nA should be as close as possible to zero matrix"*. However, if the pairwise matrix is inconsistent, Corollary 2 says that *"there must be some inconsistent elements in induced matrix C deviating far away from zero"*. By this corollary, the largest value in matrix C can be used to identify the most inconsistent element in the original matrix A. Some of the identification processes are presented next.

The procedures of the IBM model proposed in Ergu et al. [8] include three major steps, containing seven specific steps (Details are referred to Ergu et al. [8]). The first five steps are easy to be implemented by MATLAB software in practice, i.e. 1) Construct an induced matrix C=AA-nA; 2) Identify the largest absolute value(s) of elements and record the corresponding row and column; 3) Construct the row vector and column vector using the recorded location; 4) Calculate the scalar product f of the vectors; 5) Compute the deviation elements between a_{ij}

and vectors *f*. However, for Steps 6-7, the definitions are not easy to be quantified and it needs the decision makers to determine when we should use *Method of Maximum*,*Method of Minimum*, and *Method for adjusting* a_{ij} to identify the most inconsistent entries. In the following section, we combine these identification methods to perform the simulation experiment by generating randomly reciprocal matrix in order to validate the effectiveness of the induced bias matrix model.

3 Simulation experiments

3.1 Design of simulation experiments

The simulation experiments were performed to confirm the effectiveness of the induced bias matrix model using random inconsistent reciprocal matrices. We generated randomly 10^6 set of reciprocal matrices with orders 3 to 9, and 10^5 set of reciprocal matrices with orders 10-12, i.e. the entries above the main diagonal of a reciprocal matrix is generated randomly from the 17 numbers $(1/9, 1/8, 1/7, \ldots, 1, 2, 3, \ldots, 9)$ in order to satisfy the Saaty's fundamental 9-point scales, the entries below the main diagonal of the PCM is filled automatically with the corresponding reciprocal value. Then calculating the consistency ratio by the formula proposed by Saaty [7], where λ_{max} is the maximum eigenvalue of matrix A, and n is the order of matrix A. If the $CR < 0.1$, discard the generated matrix, if the $CR \geq 0.1$, then applying the IBM model to modify the inconsistent entry and improve the consistency ratio by the six steps and the combined algorithm, as shown in Figure 1. If the consistency ratio of the generated randomly reciprocal pairwise comparison matrix cannot be reduced to be lower than 0.1, then record the corresponding matrix. The specific procedures of this simulation experiment are shown in Figure 1.

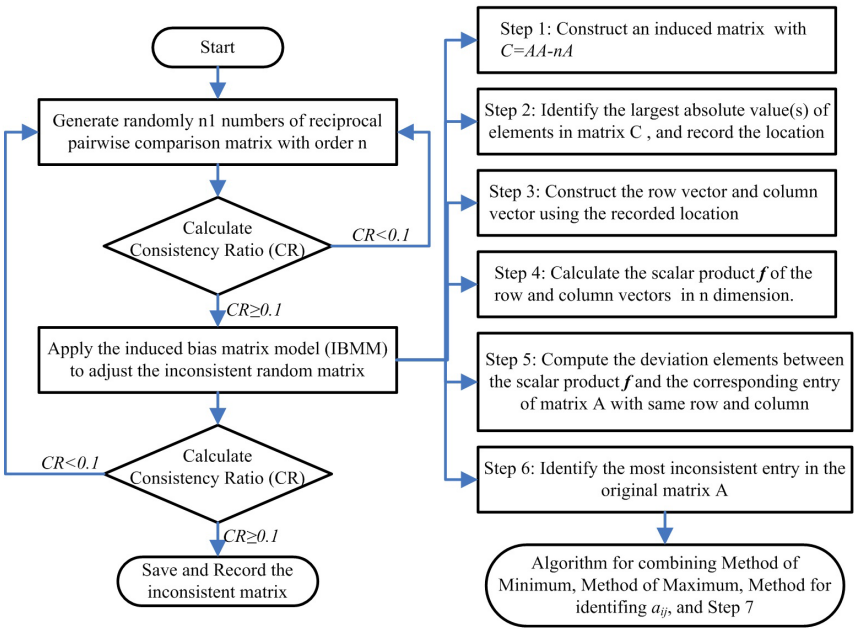


Figure 1: Procedures of simulation experiment by IBM model

For the matrices with $CR \geq 0.1$, the formula of steps 1-6 presented in Ergu et al. [8] are directly used to identify and modify the most inconsistent entries in matrix A as well as improving the consistency ratio. However, the *Method of Maximum, Method of Minimum and Method for*

identifying a_{ij} proposed in Step 6 and Step 7 involve qualitative observation and judgment, for instance, how many absolute values in vector f can be regarded as more absolute values that are around zero? How to measure the absolute values of a_{ij}, a_{ik} and a_{kj} are too large or too small by quantifying terms? Therefore, the following Algorithm is used to combine the previously mentioned identification processes.

3.2 Algorithms of simulation experiments

In order to simulate the induced bias matrix model, the program codes with two input parameters, n and n1 were written by Matlab 7.0 to randomly generate reciprocal matrices, in which n denotes the numbers of random reciprocal matrix, while n1 represents the number of simulation. For the space limitations, we omitted the first five steps, and the following Algorithm is used to combine the *Method of Maximum*,*Method of Minimum*, and *Method for adjusting* a_{ij}.

Algorithm 1: Improving the consistency ratios of the random reciprocal matrices with $CR \geq 0.1$
Input: Matrix Order, n; Number of simulation, n1
Output: Matrices with $CR \geq 0.1$
Process:
01. C=AA-nA % Matrix A is the generated randomly reciprocal matrix with $CR \geq 0.1$
02. If $c_{ij} < 0$
03. Adjust a_{ij} using $a_{ij} = \frac{1}{n-2}\sum_{k=1}^{n} a_{ik}a_{kj}$
04. End % Method for identifying $a_{ij}(1)$
05. If $c_{ij} > 0$ & & $min(f) == 0$ % We can obtain that a_{ij} is inconsistent whether it is too large or too small, in which f is the vector product
06. Adjust a_{ij} using $\widetilde{a}_{ij} = \frac{1}{n-2}\sum_{k=1}^{n} a_{ik}a_{kj}$
07. End % Method for identifying $a_{ij}(2)$
08. If $c_{ij} > 0$ % a_{ij} and a_{ik} (or a_{kj}) might have problematic
09. $[m, k] = max(f)$; % m is the element with the largest value in vector f, while k is the corresponding location.
10. If $c_{ik} < 0$ & & $c_{kj} \geq 0$ % a_{ik} is problematic (too large).
11. Adjust c_{ik} using $a_{ik} + c_{ik}/(n-2)$.
12. Break
13. End
14. If $c_{ik} \geq 0$& & $c_{kj} < 0$ % a_{kj} is problematic and large
15. Adjust a_{kj} using $a_{kj} + c_{kj}/(n-2)$
16. Break
17. End
19. If $c_{ik} < 0$ & & $c_{kj} < 0$ % c_{ik} and c_{kj} are problematic
20. If $abs(c_{ik}) >= abs(c_{kj})$
21. Adjust a_{ik} using $a_{ik} + c_{ik}/(n-2)$
22. Break
23. Else
24. Adjust a_{kj} using $a_{kj} + c_{kj}/(n-2)$
25. Break
26. End
27. End

28. If $f(k) > 0$ & & $c_{ik} >= 0$ & & $c_{kj} >= 0$ % It is unreasonable to occur simultaneously, if it does occur, then go to adjust the second largest value.
29. $c_{ij} = 0$;
30. End
31. End
32. Calculate the CR; % see Algorithm 2
33. If $CR < 0.1$
34. Break
35. End.

Algorithm 2: Calculating the consistency ratio of the modified matrix B
Input: Modified random matrix B
Output: Consistency ratio CR.
Process:
01. n=length(B); % B is reciprocal matrix
02. RI=[1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 0 0 0.52 0.89 1.12 1.26 1.36 1.41 1.46 1.49 1.52 1.54 1.56 1.58 1.59];
03. [a,b]=eig(B);
04. [c1,d1]=max(b); % c1 is maximum value in each column, d1 is the corresponding row of each element
05. [e1,f1]=max(c1); % e1 is the largest element, f1 is the corresponding column.
06. CI=(e1-n)/(n-1);
07. CR=CI/RI(2,n);

3.3 Experimental results

In this section, we do not attempt to optimize the program codes for speed, therefore, we set the matrix order n to be 3-12, and the simulation number n1=10^6 for the matrices with orders 3-9. For the matrix with orders 9-12, we only simulated 10^5 numbers of randomly reciprocal matrices. The results of simulation experiments are shown in Table 1. It can be seen that some of the random reciprocal matrices with orders from 3 to 6 passed the consistency test, for instance, 206130 random matrices with order 3 passed the consistency test among 10^6 matrices, while 73 random matrices with order 6 passed the consistency text. However, all random matrices with orders 7-12 did not pass the consistency test. For the random matrices with $CR \geq 0.1$, the proposed IBM model was used to modify the most consistent entries and improve the consistency ratio. Table 1 shows that the consistency ratios of all the inconsistent random matrices with $CR \geq 0.1$ and orders 3-7 have been improved and lower than 0.1 after the proposed IBM model is used to modify the random matrices, as shown in Figures 2-7, while some matrices still failed the consistency test, the numbers are 3 for order 8, 5 for order 9, 1 for order 10, 2 for order 11 and 13 for order 12, as shown in Figures 8-12. The corresponding simulation Figures for 10^6 random matrices with orders 3 to 9, and 10^5 random matrices with orders 10-12 are shown in Figures 2-10.

Table 1 Simulation experiments for randomly generated matrices with different orders

Matrix Order	Number of simulation	Number of matrices with $CR \geq 0.1$	Failed matrices	Succeeded Matrices
3	1000000	793870	0	793870
4	1000000	968083	0	968083
5	1000000	997518	0	997518
6	1000000	999927	0	999927
7	1000000	1000000	0	100000
8	1000000	1000000	3	999997
9	1000000	1000000	5	999995
10	100000	100000	1	99999
11	100000	100000	2	99998
12	100000	100000	13	99987

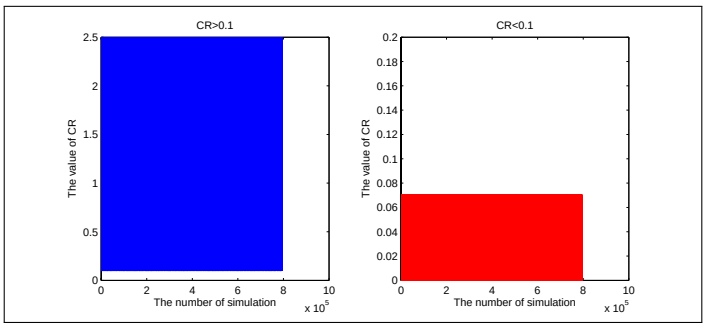


Figure 2: Simulation experiment for 10^6 randomly generated matrices with order 3

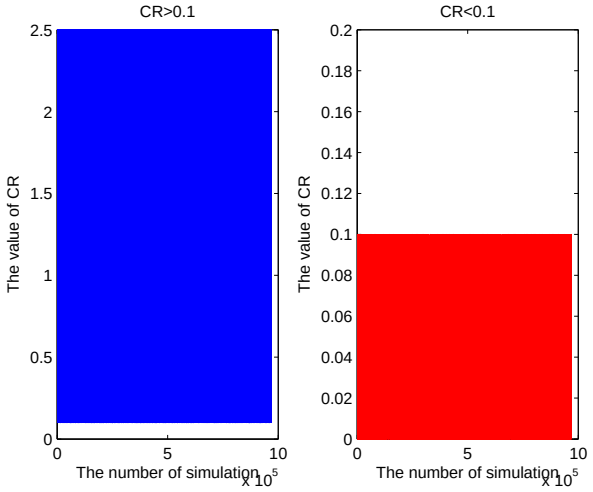


Figure 3: Simulation experiment for 10^6 randomly generated matrices with order 4

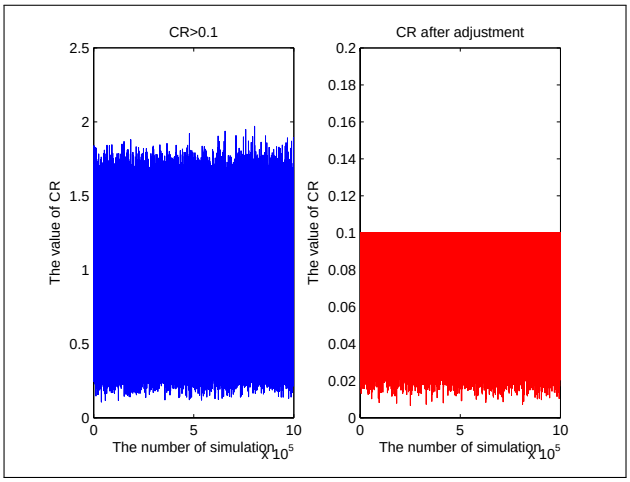


Figure 4: Simulation experiment for 10^6 randomly generated matrices with order 5

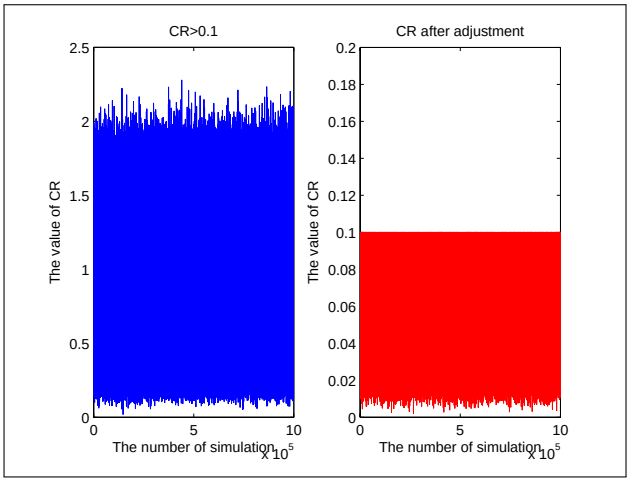


Figure 5: Simulation experiment for 10^6 randomly generated matrices with order 6

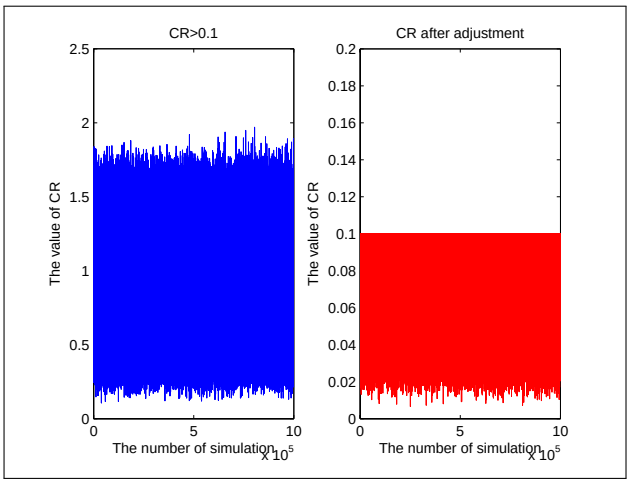


Figure 6: Simulation experiment for 10^6 randomly generated matrices with order 7

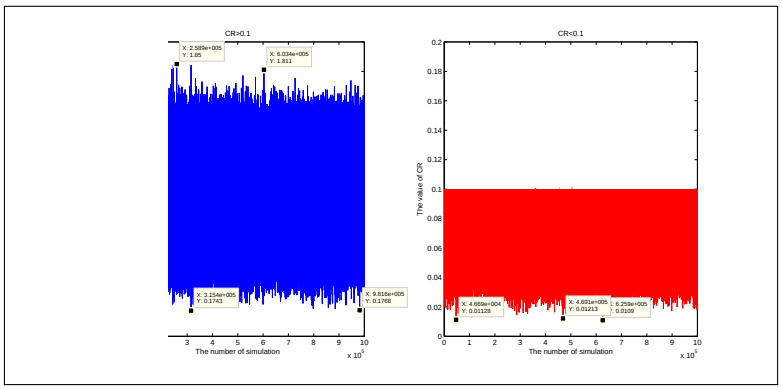


Figure 7: Simulation experiment for 10^6 randomly generated matrices with order 8

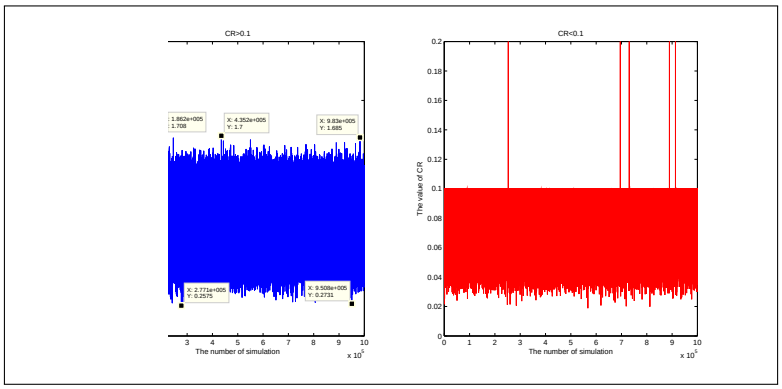


Figure 8: Simulation experiment for 10^6 randomly generated matrices with order 9

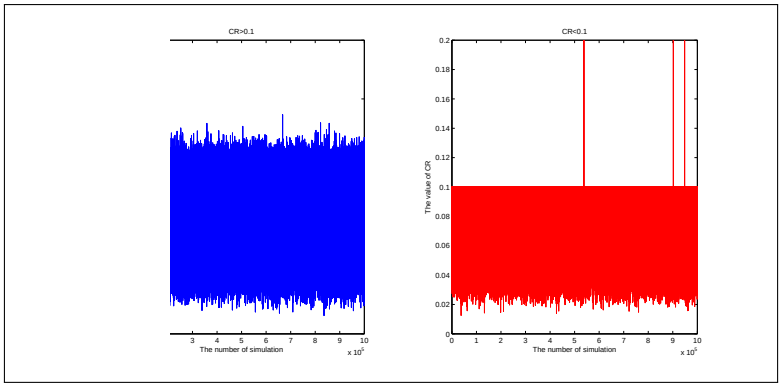


Figure 9: Simulation experiment for 10^5 randomly generated matrices with order 10

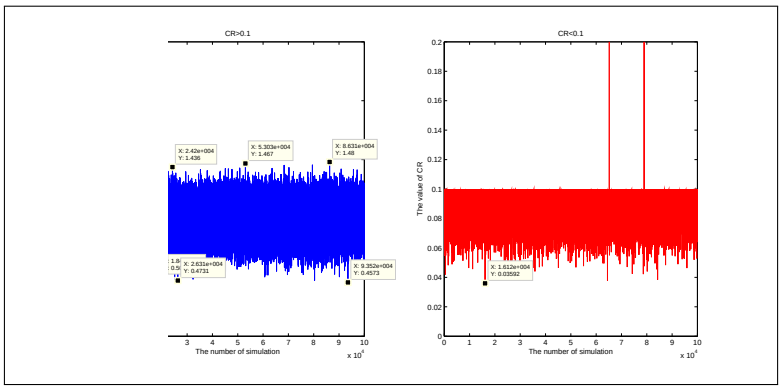


Figure 10: Simulation experiment for 10^5 randomly generated matrices with order 11

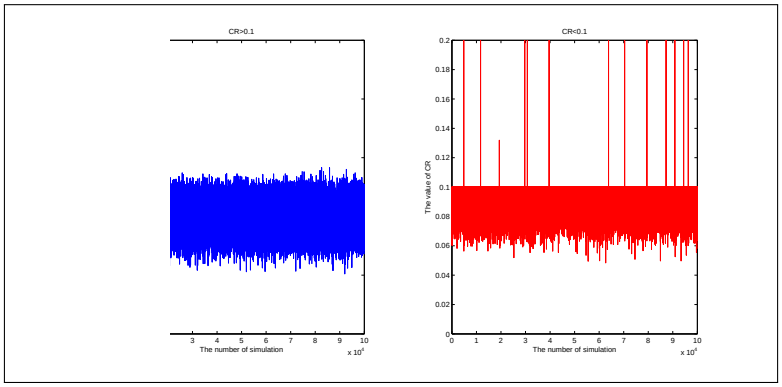


Figure 11: Simulation experiment for 10^5 randomly generated matrices with order 12

4 Conclusions

In this paper, some of the identification processes proposed in Ergu et al. [8] were combined to implement the programming. Based on these combinations, an algorithm was proposed and simulation experiments on random reciprocal matrices with different orders were conducted to validate the effectiveness of the induced bias matrix model. We found that some matrices generated randomly could pass the consistency test, and the higher the orders of matrices are, the less the matrices have $CR < 0.1$. When the orders of random matrices increase to 7, all matrices generated randomly have $CR \geq 0.1$, and they need to be adjusted. After the algorithm of the induced bias matrix (IBM) model was applied to these matrices, all the consistency ratios of random matrices with orders 3-7 were improved and less than the threshold 0.1, while fewer matrices with order higher than 8 still could not be modified satisfactorily. However, we believe that the consistency of the pairwise comparison matrices provided by experts will be much better than the consistency of random matrices, thus the proposed IBM model is capable of dealing with the consistency of a PCM.

Although the results of the simulation experiments show the effectiveness of the IBM model, the experimental findings also reveal the failed tendency will increase with the increase of the matrices order. The failed matrices should be paid more attention to and analyze the reason why it failed the consistency test, we leave it for further research in future.

Acknowledgements

This research was supported in part by grants from the National Natural Science Foundation of China (#71222108 for G. Kou, #71325001 and #71173028 for Y. Peng, #71373216 for D. Ergu, #91224001 for S.M.Li);Program for New Century Excellent Talents in University #NCET-12-0086 for Y. Peng, and the Research Fund for the Doctoral Program of Higher Education (#20120185110031 for G. Kou). We thank Prof. Waldemar W. Koczkodaj for encouraging us to conduct simulation experiments in order to further validate the effectiveness of IBM model.

Bibliography

[1] Thurstone, L. (1927), A law of comparative judgment, *Psychological review*, 34(4): 273-273.

[2] Saaty, T. L. 1986. Axiomatic Foundation of the Analytic Hierarchy Process, Management Science 32(7): 841-855.

[3] Saaty, T. L. (1994), How to Make a Decision: The Analytic Hierarchy Process, *Interfaces*, 24: 19-43.

[4] Siraj, S.; Mikhailov, L.; Keane, J. (2012), A heuristic method to rectify intransitive judgments in pairwise comparison matrices, *European Journal of Operational Research*, 216: 420-428.

[5] Kou, G., Ergu,D., Shang,J. (2014). Enhancing Data Consistency in Decision Matrix: Adapting Hadamard Model to Mitigate Judgment Contradiction, *European Journal of Operational Research*, 236(1):261-271.

[6] Saaty, T. L. (1980), *The Analytical Hierarchy Process*, New York: McGraw-Hill.

[7] Saaty, T. L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary, *European Journal of Operational Research*, 145(1): 85-91.

[8] Ergu, D. ; Kou, G.; Peng, Y.; Shi, Y. (2011). A Simple Method to Improve the Consistency Ratio of the Pair-wise Comparison Matrix in ANP, *European Journal of Operational Research*, 213(1): 246-259.

[9] Harker, P. T. 1987. Derivatives of the Perron root of a positive reciprocal matrix: With applications to the analytic hierarchy process, *Applied Mathematics and Computation*, 22: 217-232.

[10] Xu, Z., Wei, C. (1999),A consistency improving method in the analytic hierarchy process, *European Journal of Operational Research*, 116: 443-449.

[11] Cao, D., Leung, L. C., Law, J. S. (2008). Modifying inconsistent comparison matrix in analytic hierarchy process: A heuristic approach, *Decision Support Systems*, 44: 944-953.

[12] Choo, E., Wedley, W. (2004). A common framework for deriving preference values from pairwise comparison matrices, *Computer and Operations Research*, 31 (6): 893-908.

[13] Lin, C. (2007). A revised framework for deriving preference values from pairwise comparison matrices, *European Journal of Operational Research*, 176 (2): 1145-1150.

[14] Kou,G, Lin,C. (2014). A cosine maximization method for the priority vector derivation in AHP, *European Journal of Operational Research*, 235:225-232.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):419-429, August, 2014.

Determining the State of the Sensor Nodes Based on Fuzzy Theory in WSNs

M.S. Gharajeh

Mohammad Samadi Gharajeh
Department of Computer Engineering
Tabriz Branch, Islamic Azad University
Tabriz, Iran
mhm.samadi@gmail.com

Abstract: The low-cost, limited-energy, and large-scale sensor nodes organize wireless sensor networks (WSNs). Sleep scheduling algorithms are introduced in these networks to reduce the energy consumption of the nodes in order to enhance the network lifetime. In this paper, a novel fuzzy method called Fuzzy Active Sleep (FAS) is proposed to activate the appropriate nodes of WSNs. It uses the selection probability of nodes based on their remaining energy and number of previous active state. The proposed method focuses on a balanced sleep scheduling in order to belong the network lifetime. Simulation results show that the proposed method is more efficient and effective than the compared methods in terms of average network remaining energy, number of nodes still alive, number of active state, and network lifetime.
Keywords: wireless sensor networks (WSNs), fuzzy theory, sleep scheduling, energy consumption, network lifetime.

1 Introduction

Wireless Sensor Networks (WSNs) organize a wireless network, where the sensors called nodes have basically some features such as low cost, limited energy, and large scalability [1]. A very large number of nodes are deployed in the network to sense and to transmit the environmental conditions or occurrence of the physical events. They are used in some applications such as traffic management, close-circuit camera in retail, seismic monitoring, and military usage. Organization of the nodes in a hierarchical topology, short length messages, short range of message transfer, and sleep scheduling of the nodes are some of the essential mechanisms to reduce the energy consumption of nodes in order to enhance the network lifetime [2].

The sleep scheduling categorizes the nodes into active or sleep states. A proper sleep scheduling method performs the scheduling operation in a way that maintains the connectivity among nodes and coverage of whole network. A sensor network is connected when each active node can transmit its data via one or multiple hops toward a specified center. Meanwhile, the coverage is specified as an area that can be accessed by active nodes. Both of the connectivity and coverage are the essential factors to monitor a given area that should be considered by the presented sleep scheduling methods.

Fuzzy logic is utilized to develop some models such as physical tools, uncertain and complex systems, and non-linear processes. The fuzzy models can be easily understood, and have less external complexity with more useful features [3]. Furthermore, fuzzy controllers can take an appropriate decision even by imprecise and incomplete information. The linguistic terms and inexact data can be manipulated as a useful tool to design the uncertain systems. Therefore, the sleep scheduling can be designed and implemented by fuzzy decision making to give most advantageous in terms of connectivity, coverage, and network lifetime.

The rest of this paper is organized as follows. The related works presented in section 2 discuss about some of the prior sleep scheduling methods. Section 3 describes the proposed fuzzy method by addressing the designated fuzzy system. Performance evaluation of the simulated methods is explained in section 4. Finally, the paper is concluded in section 5.

2 Related Works

A sleep scheduling method for stationary nodes of WSNs is developed in [4] that utilize a discrete-time Markov chain (DTMC) model. It applies an analytical method using slotting the time according to the data unit transmission time in order to discover the trade-offs existing between energy conserving and system throughout metrics including network capacity, energy consumption, and data delivery ratio. Besides, the sensor nodes are considered as three operational modes as transmit, receive, and idle to simply adapt them with various traffic conditions.

An analytical method for the random scheduling algorithms is provided in [5] to derive the detection probability and detection delay. The simulation results are carried out with discrete event simulation to investigate the impact of number of subsets and number of sensor nodes on coverage intensity, detection probability, and detection delay. A random scheme for WSNs is also proposed in [6] that develop an analytical schema to investigate the relation between randomized sleep state and network throughout. A queue model for nodes and an efficiency framework for whole the network are included in the presented framework to derive the throughout, energy consumption, and delay time. Another randomized scheduling method is studied in [7] via analysis and simulations in aspects of detection probability, detection delay, and network coverage intensity. Furthermore, a problem of prolonging the network lifetime under Quality of Service (QoS) limitation such as bounded detection probability, detection delay, and network coverage intensity is analyzed by authors.

An optimal sleep control mechanism is proposed in [8] to prolong the network lifetime with reducing the energy consumption of the nodes. It utilizes the proposed procedure by distance between the sensor nodes and the sink. Furthermore, energy of whole the network is balanced through reducing the number of transmissions related to the sensor nodes which are placed more close to the sink. In the method presented in [9], several characteristics of active/sleep model in WSNs are investigated. The main mechanism of this method to manage the nodes in an ON or OFF period is that the steady-state probability distribution of number of packets is derived in the reference node. Another method is presented in [10] that determine the active and sleep modes of the nodes as randomly or alternatively manner in a stochastic model of WSNs. The active mode is categorized as two phases as full active phase and semi-active phase to better manage the energy consumption of the nodes. This method evaluates the energy consumption of the network by developing the important analytical formulae.

3 The Proposed Method

The main objective of the proposed method is to balance the energy consumption of the nodes in order to enhance the network lifetime. They are possible by activating and sleeping the appropriate nodes for a period time based on fuzzy decision making. If the nodes are activated based on a structural scheduling method, their energy are consumed in an equivalent flow so that the energy balance of the network will be considerably enhanced. If various types of the sensors such as temperature, smoke, and light intensity are used in the network, the proposed method will be applied within each category independently. The proposed fuzzy method called Fuzzy Active Sleep (FAS) operates based on fuzzy decision making. The whole network is divided to different areas so that only one node is activated in each area for a period time. It is worth noting that the number of divided areas is determined based on network size and number of nodes. A single sink is assumed in the centre point of network that determines the active nodes of the areas. Furthermore, it receives the environmental data from nodes, and forwards them to the base station. The reason is that all the nodes cannot directly transmit their data to the base station. An overview of the assumed network is shown in Figure 1, so that it is composed

of four areas included by three types of sensors as temperature, smoke, and light intensity.

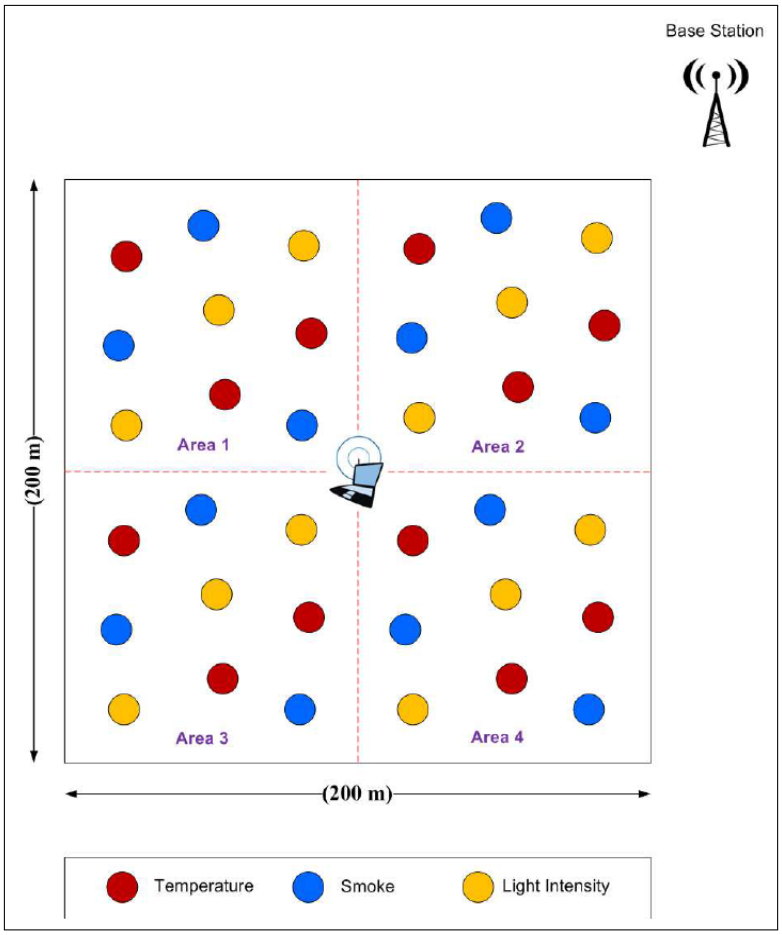


Figure 1: An overview of the assumed network

There are some input and output variables in the fuzzy systems to make the fuzzy rules. The fuzzy rules are used in a fuzzy system to decide an appropriate action in the uncertain conditions. The input variables of the proposed fuzzy system are determined as follow; the first one is remaining energy of nodes (denoted by REi); the second one is the number of previous active state (denoted by ASi); here, "i" refers to the number of existing nodes in the related area. The output variable is selection priority of the nodes (specified by SPi). Suppose for this work, REi and SPi take on following linguistic values: VL (very low), L (low), M (medium), H (high), VH (very high); and ASi take on the following linguistic values: FE (feeble), FW (few), ME (medium), MA (many), L (lots). Membership graph for the inputs and the output variables are depicted in Figure 2. Note that the membership functions of ASi are defined by Triangular [11] method, and membership functions of REi and SPi are determined by Bell-shaped [11] method. While there are two input variables as each one can accept five linguistic terms, total number of the fuzzy rules is $5^2 = 25$. Some of the fuzzy rules used in the proposed fuzzy system are represented in Table 1. Note that the fuzzy rules are constructed by Mamdani-type fuzzy rule-based systems [12]. Meanwhile, all the fuzzy rules are aggregated together by OR operator to produce the total fuzzy rule.

A Schematic of the fuzzy rules used in the proposed fuzzy system is shown in Figure 3 based on input and output variables.

The appropriate node from among the nodes' groups within each area is selected by sink

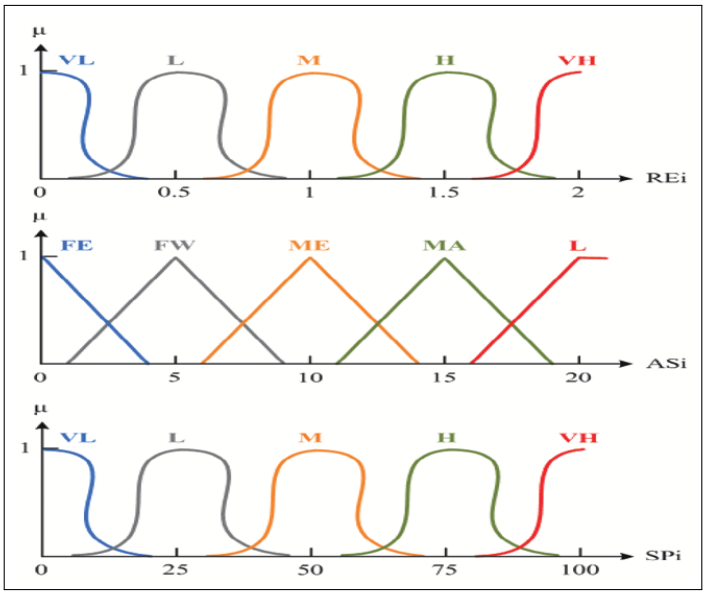
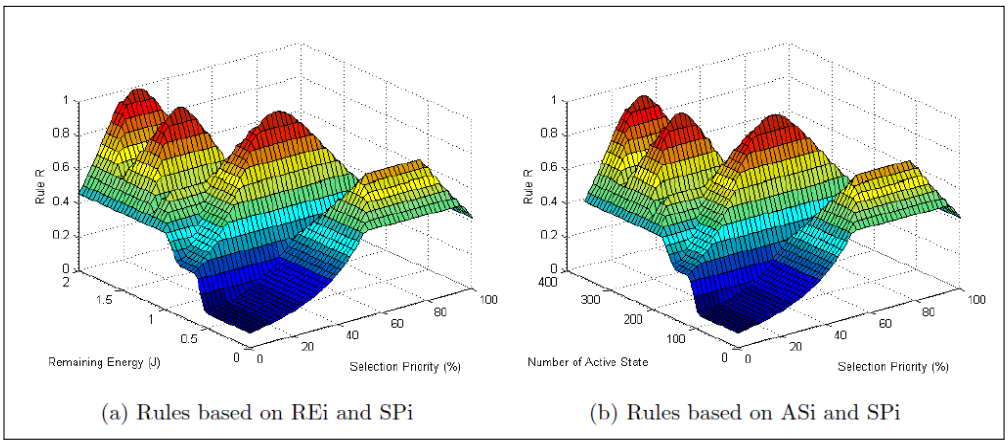


Figure 2: Membership graph for the inputs (remaining energy and number of active state) and the output (selection priority)

Table 1: Some of the fuzzy rules used in the proposed fuzzy system

Rule No.	Input variables		Output variable
	REi	ASi	SPi
1	VH	FE	VH
2	H	FW	H
3	M	ME	M
4	VH	L	L
5	H	MA	L
6	L	L	L
7	VL	FW	H
8	L	FE	H
9	M	MA	L
10	VH	ME	VL



(a) Rules based on REi and SPi (b) Rules based on ASi and SPi

Figure 3: A schematic of the fuzzy rules used in the proposed fuzzy system

using the proposed method for a period time as described following. First, remaining energy of the nodes is converted to a fuzzy value by Bell-shaped membership function, and the number of active state at nodes is converted by Triangular membership function. Then, the fuzzy value of each node's selection priority is determined by the approximate reasoning, total fuzzy rule, and the inputs' fuzzy values. Finally, the output fuzzy value is converted to a crisp value by center-of-gravity [13] which is a defuzzification method as follows

$$SPi = \frac{\sum_{i=1}^{n} \mu_{sp}(x_i).x_i}{\sum_{i=1}^{n} \mu_{sp}(x_i)} \tag{1}$$

Where n indicates the number of elements in the universe set of selection priority, x_i represents each elements of the universe set, and $\mu_{sp}(xi)$ describes the membership degree of x_i in the universe set. The node with highest crisp value is selected as active node, and other nodes are selected as sleep nodes.

4 Performance Evaluation

The simulation processes are carried out in MATLAB. 40 sensor nodes are randomly deployed in a topographical area of dimension 200 m × 200 m. All nodes have the same initial energy 2J. The proposed method is compared to All Active method and Random Active Sleep method called "RAS" to evaluate them in terms of average network remaining energy, number of nodes still alive, and number of active state. Furthermore, impacts of the different experimental parameters such as interval time between data sense, interval time between sending data, and initial energy of nodes on the network lifetime are evaluated carefully. Note that all the nodes are always activated in the All Active method, and the active node is selected by a random procedure in the RAS. The active nodes transmit their environmental data to the sink in a specific interval time. Afterwards, the sink also transmits the aggregated data to the base station in a determined interval time. Note that the gathered data are aggregated by the sink as follows

$$d_{Agg} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{2}$$

Where n indicates the number of data presented in the sink's buffer and d_i refers to each data of the buffer. The simulation will be terminated when the remaining energy of all the nodes is under threshold energy. Note that the discrete simulation results are the average value of the results which are independently simulated for 10 times. The transmission and receiving energies are calculated based on the model expressed in [14]. According to this mode, for transmitting an l-bit data packet a long a distance d, the radio spends

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2 , & d < d_0 \\ lE_{elec} + l\epsilon_{mp}d^4, & d \geq d_0 \end{cases} \tag{3}$$

Where d_0 indicates a threshold distance, $\epsilon_{fs}d^2$ uses the free space (fs) model to calculate the amplifier energy, and $\epsilon_{mp}d^4$ utilizes the multipath (mp) model to estimate the amplifier energy. Meanwhile, the spending radio to receive this data packet is calculates as

$$E_{Rx}(l) = lE_{elec} \tag{4}$$

Note that the threshold energy to sense or to receive an l-bit data packet is calculated like $E_{Rx}(l)$. The simulation parameters and their default values are represented in Table 2.

Table 2: Simulation parameters

Parameter	Value
Topographical area (meters)	200 × 200
Sink location (meters)	(100, 100)
Location of base station (meters)	(500, 500)
Buffer size of the sink (Packet)	10,000
Buffer size of the base station (Packet)	10,000
Number of nodes	40
Initial energy of node	2 J
Interval time between data sense (Round)	10
Interval time between sending data (Round)	5
Interval time between active sleep changes (Round)	25
Interval time between data transmission through the sink (Round)	20
E_{elec}	50 nJ/bit
ϵ_{fs}	10 pJ/bit/m$\hat{2}$
ϵ_{mp}	0.0013 pJ/bit/m$\hat{4}$

4.1 An Instance of Selecting the Active Node by the Proposed Fuzzy System

As previously described, the active node is independently selected within each area. Selection priority is calculated for all the nodes presented in the areas; then, the node with the highest priority is activated for a period time. As represented in Table 3, if there are five nodes in a special area, selection priority is calculated based on their remaining energy and number of previous active state. Therefore, the node N4 which has the highest priority is selected as the active node for a period time.

Table 3: An example of determining the selection priority for each node

	Input variables		Output variable
Node	REi	ASi	SPi
N1	2	11	44.913
N2	1.2	15	42.027
N3	0.3	5	50.843
N4	0.8	2	54.757
N5	1.6	8	48.058

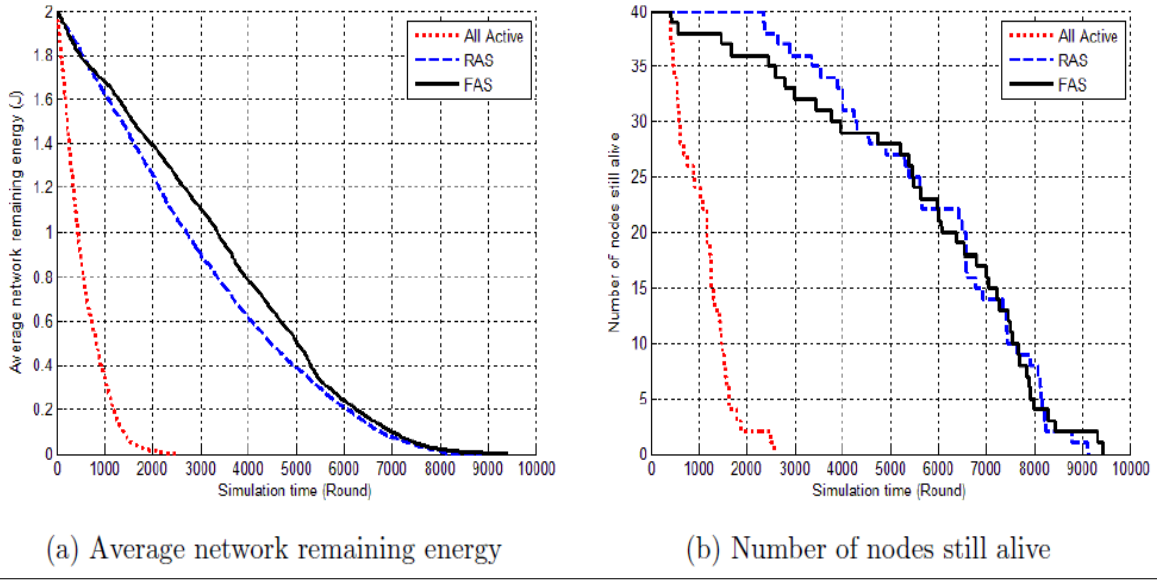
4.2 Simulation Results

Some of the continuous simulation results of the evaluated methods are shown in Figure 4 in terms of average network remaining energy and number of nodes still alive. These parameters illustrate the lifetime status of network under the situations represented in Table 3. As shown in the results, the network lifetime obtained by the All Active method is 2,560 Round, by the RAS method is 9,125, and by the proposed method is 9,440. As it has been expected, the lifetime in the All Active is very low; but it is near to each other in the RAS and FAS methods. However, the average network energy and number of live nodes in the proposed method is higher than that of RAS due to balance the active state of the nodes. Average network remaining energy are

calculated in each round as follows

$$Rem_{Avg} = \frac{\sum_{i=1}^{n} Rem_e(i)}{n} \tag{5}$$

Where n indicates the number of nodes and $Rem_e(i)$ represents the remaining energy at each node "i". Simulation results demonstrate that the average network remaining energy and the number of nodes still alive achieved by the proposed method could be increased by about 700% more than that obtained by the All Active method and by about 10% more than that obtained by the RAS method.



(a) Average network remaining energy (b) Number of nodes still alive

Figure 4: Lifetime status of the network in a simulation execution based on various methods

The number of dead nodes is calculated according to Algorithm 1 in each round. If it equals to the number of nodes, the simulation process will be terminated. Note that the number of last round is known as network lifetime.

Algorithm 1: Calculate the number of dead nodes

```
n ⟵number of nodes;
count ⟵0;
i ⟵1;
l ⟵data packet size (bit);
E_elec ⟵50×10^-9;
E_th ⟵1×E_elec;
while i ≤n do
    if Rem_e(i) < E_th then
        count ⟵count+1;
return count;
```

Number of active state at nodes is one of the most important factors in the sleep scheduling methods. A good method tries to balance this factor in order to enhance the network lifetime. As shown in Figure 5, the number of active state in the All Active method has a stationary

high value due to activate all the nodes in all the active/sleep selection rounds. Moreover, this value in the proposed method is more balance than that of the RAS method. The reason is that selecting the active and sleep nodes is determined based on remaining energy of nodes and number of previous active state that leads to balance the final number of active state. Note that the number of active state at each node increases when it is selected as the active node for a period time.

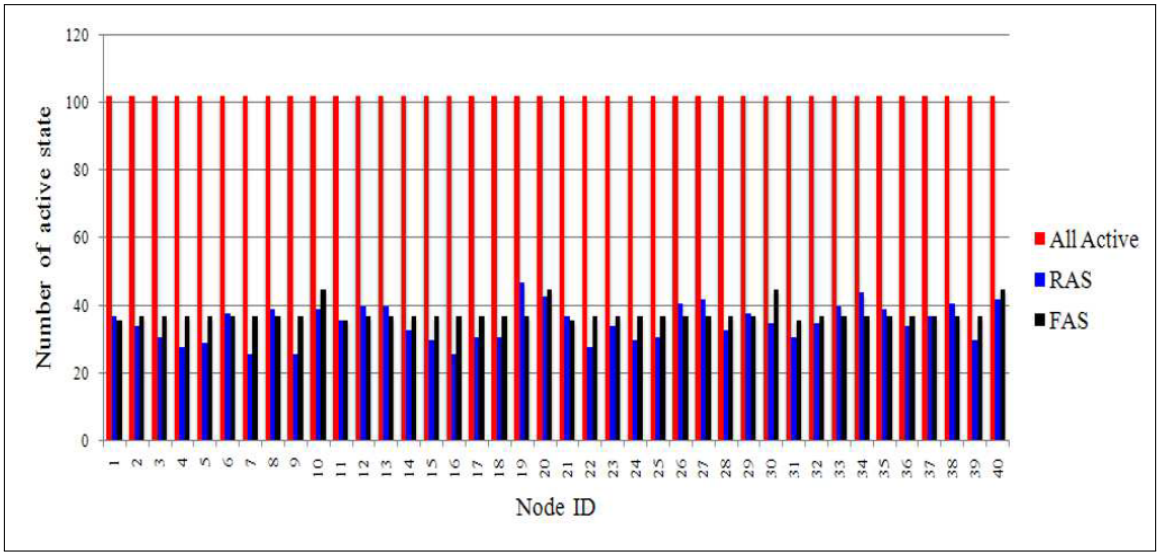


Figure 5: Number of active state at nodes under various methods

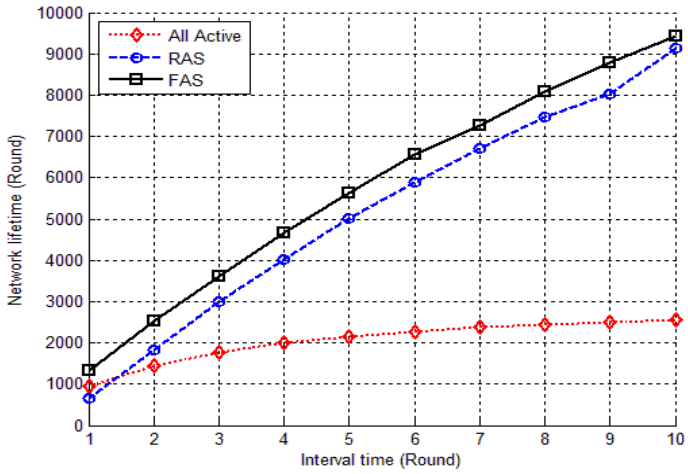
The statistical results of the active state at nodes are represented in Table 4. The four famous statistical functions including Minimum, Maximum, Mode, and Standard Deviation are used to determine the detailed information. The Minimum and Maximum functions specify the range of all the numbers sets. The Mode function represents the most frequently occurring, or repetitive, number of active state at all the numbers sets. Meanwhile, the Standard Deviation function specifies a measure of how widely numbers are dispersed from the average number so that it can be calculated as follows

$$Std_{dev} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{6}$$

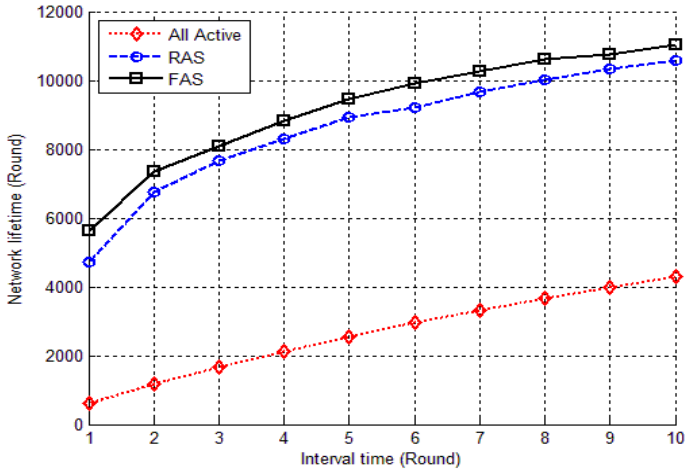
Where n indicates the number of nodes, x_i represents the number of active state at node "i", and $\bar{x}$ specifies the average value of all the numbers. The values calculated by various statistical functions represents that the proposed method is more efficient and balance than both of the other methods.

Table 4: The statistical values of the active state at nodes under various methods

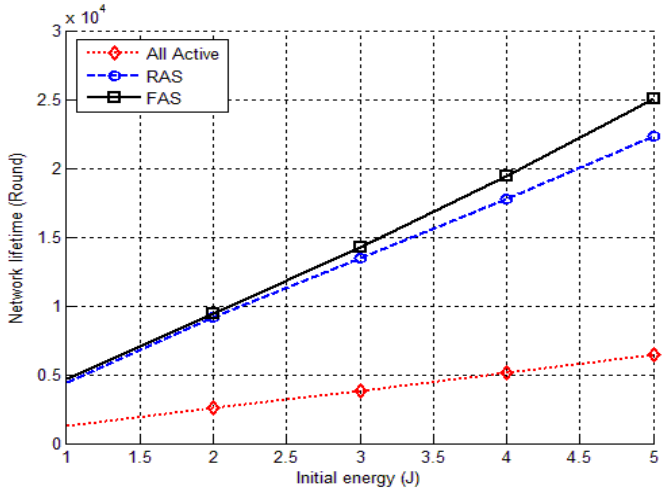
	Method		
Statistical function	All Active	RAS	FAS
Minimum	120	26	36
Maximum	120	47	45
Mode	120	31	37
Standard Deviation	0	5.44	2.48



(a) Interval time between data sense



(b) Interval time between sending data



(c) Initial energy

Figure 6: Network lifetime achieved by different methods under various parameters

Some parameters such as interval time between data sense, interval time between sending data, and initial energy influences strongly on the network lifetime. When each of these parameters increases, the network lifetime will be considerably enhanced. The interval time between data sense determines a period time to sense the environmental conditions by nodes. Meanwhile, the interval time between sending data specifies a period time to transmit the sensed data to the sink. Affection of them on the various methods is illustrated in Figure 6. As shown in the results, the All Active method has the network lifetime very lower than others. Besides, the proposed method surpasses the RAS method under various parameters changes. The reason is that some of the nodes are activated more than other nodes in the RAS method so that the energy consumption of the nodes is unbalanced. The unbalanced energy consumption of nodes causes the network lifetime to be lower than the proposed method.

In addition, the energy efficiency of the network can be specified based on total energy consumption of the nodes which is calculated as follows

$$T_e = \sum_{i=1}^{n} Cons_e(i) \tag{7}$$

Where n indicates the number of nodes and $Cons_e(i)$ represents the energy consumption of each node "i". Note that the above formulae can be used to determine the energy consumption of the nodes both in each round and whole the network.

5 Conclusions

Wireless Sensor Networks (WSNs) are composed of some large-scale, low-cost, and limited-energy sensor nodes. The sleep scheduling methods presented in the WSNs cause the network lifetime to be considerably enhanced. In this paper, a novel fuzzy method called Fuzzy Active Sleep (FAS) proposed to select the appropriate node in each desired area to be activated for a period time. It selects the active node from among the related nodes based on their remaining energy and number of previous active state. Selection procedure of the active nodes is balanced by the proposed method that leads the network lifetime to be enhanced. Simulation results represent that the proposed method surpasses the other compared methods in aspects of average network remaining energy, number of nodes still alive, number of active state, and network lifetime.

Bibliography

[1] Stankovic, J.A. (2008); Wireless sensor networks, *Computer*, ISSN 0018-9162, 41(10): 92-95.

[2] Zhang, P. et al (2013); Clustering algorithms for maximizing the lifetime of wireless sensor networks with energy-harvesting sensors, *Comput Netw*, ISSN 1389-1286, 57(14): 2689-2704.

[3] Cerami, M.; Straccia, U. (2013); On the (un)decidability of fuzzy description logics under Lukasiewicz t-norm, *Inform Sciences*, ISSN 0020-0255, 227: 1-21.

[4] Chiasserini, C.F.; Garetto, M. (2004); Modeling the performance of wireless sensor networks, *Proc. IEEE Infocom ser*, ISSN 0743-166X, 220-231.

[5] Xiao, Y. et al (2007); Modeling detection metrics in randomized scheduling algorithm in wireless sensor networks, *Proc. IEEE WCNC*, ISSN 1525-3511, 3741-3745.

[6] Liu, J. et al (2010); Analysis of random sleep scheme for wireless sensor networks, *International Journal of Sensor Networks*, ISSN 1748-1279, 7(1): 71-84.

[7] Xiao, Y. et al (2010); Coverage and detection of a randomized scheduling algorithm in wireless sensor networks, *IEEE T Comput*, ISSN 0018-9340, 59(4): 507-521.

[8] Keh, H.C. et al (2011); Power saving mechanism with optimal sleep control in wireless sensor networks, *Tamkang J. Sci. Eng*, ISSN 1560-6686, 14(3): 235-243.

[9] Li, W.W. (2011); Several characteristics of active/sleep model in wireless sensor networks, *Proc. IEEE NTMS'4*, ISSN 2157-4952, 1-5.

[10] Zhang, Y.; Li, W. (2012); Modeling and energy consumption evaluation of a stochastic wireless sensor network, *Eurasip J Wirel Comm*, ISSN 1687-1499, 2012(1): 1-11.

[11] Zhao, J.; Bose, B.K. (2002); Evaluation of membership functions for fuzzy logic controlled induction motor drive, *Proc. IEEE IECON'2*, 1: 229-234.

[12] Alcalá, R. et al (1999); Approximate Mamdani-type Fuzzy Rule-Based Systems: Features and Taxonomy of Learning Methods, *Citeseer*, Technical Report DECSAI-990117, pp.1-23.

[13] Runkler, T.A. (1997); Selection of appropriate defuzzification methods using application specific properties, *IEEE T Fuzzy Syst*, ISSN 1063-6706, 5(1): 72-79.

[14] Heinzelman, W.B. et al (2002); An application-specific protocol architecture for wireless microsensor networks, *IEEE T Wirel Commun*, ISSN 1536-1276, 1(4): 660-670.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):430-438, August, 2014.

Distributed Compressed Sensing Algorithm for Hierarchical WSNs

N. Jiang, H. You, F. Jiang, L. Liu, Y. He

Nan Jiang*
College of Information Engineering, East China Jiaotong University
Nanchang, 330013, JiangXi, P.R.China
State Key Laboratory Breeding Base of Nuclear Resources and Environment,
East China Institute of Technology, Nanchang, 330013, JiangXi, P.R. China
*Corresponding author: jiangnan1018@gmail.com

Hui You, Lingfeng Liu
College of Information Engineering, East China Jiaotong University
Nanchang, 330013, JiangXi, P.R.China
youhui106@gmail.com, liulingfeng1983@gmail.com

Feng Jiang
School of Computer Science and technology, Harbin Institute of Technology
Harbin, 15001, HeiLongJiang, P.R.China
fjiang@hit.edu.cn

Yueshun He
College of Information Engineering, East China Institute of Technology
Nanchang, 330013, JiangXi, P.R.China
hys8418@163.com

Abstract: In the traditional theory of CS, each sensor node in Wireless Sensor Networks (WSNs) sends the information to sink node directly, and only considers the correlation between internal nodes information when recover, this would lead to the loss of the node information and also cause too much energy consumption. In this paper, combined with DEEC protocol, based on Simultaneous orthogonal matching pursuit (SOMP) algorithm, we propose a novel distributed compressed sensing algorithm for Hierarchical Wireless Sensor Networks (DCSH algorithm). This algorithm use the spatial correlation between the nodes and joint sparse model JSM-2, to compress and recover the sensor node information according to SOMP algorithm. Simulation results show that DCSH algorithm can not only obtain accurate reconstruction values of the node information and reduce the energy consumption greatly, but also can prolong the network lifetime.
Keywords: Distributed Compressed Sensing; DEEC algorithm; JMS-2 model; Cluster Architecture; Wireless Sensor Networks (WSNs).

1 Introduction

In the past few years, the capability of receiving and transmitting data in wireless sensor networks has been constantly enhanced, simultaneously, the amount of data, which need to be handled, is also growing very quickly. Traditional Nyquist sampling theory, needs the sampling rate of signal not less than 2 times the bandwidth of the signal; undoubtedly, this is higher requirement for signal processing capability and hardware equipments. Therefore, How to deal with bigger and faster data and find a new method is attracted more and more attentions. In 2004, Donoho and Candes et al proposed Compressed Sensing (CS) theory, it is a novel theory which make full use of sparse and compressibility of the signal [1] [2]. This theory shows that when the signal is sparse or compressible, it can realize the refactoring value of signal exactly

through collecting the projection values of small amount of signal. Different from traditional data processing method, sampling and compression of signal can conduct at the same time with a low rate, to the method of data processing, sampling process also completed the compression, so the amount of sampling data is greatly reduced, the Nyquist sampling is evenly spaced sampling, and compressed sampling is random sampling.

Generally, compressed sensing theory study how to use the intra-signal on compression. Considering distributed deployment and limited capability of sensor nodes in wireless sensor networks, it is necessary to use inter-signal with Distributed Compressed Sensing (DCS), D.Baron gave the basic concept and theory of Distributed Compressed Sensing [3], and then proved the upper and lower bounds of the number of measurements required for decoding [4]. In [5], the author presented another joint sparse model for the application scenarios such as the MIMO communication and speech signal, and designed the corresponding joint decoding algorithm. However, data fusion technology for wireless sensor networks based on the DCS theory is still at the starting stage. Therefore, it is necessary to study how DCS theory do the joint reconstruction through the observation vectors of each node in cooperation way, through define joint sparse of each node based on spatial correlation data.

Our contributions are showed as follows:

- Based on JSM-2 model, signals are related to each other, each signal has different coefficient value, while they are made up with the base vector of the same sparse coefficient set.
- By using SOMP algorithm, the decoding process can be simplified and corresponding measured values can also be reduced. In decoder side, the reconstruction efficiency will be improved.
- Cluster architecture is widely used in WSNs, such as classic LEACH, HEED, DEEC and so on. We try to bring cluster into CS theory. On one hand, it will save the transmission energy; on the other hand, sink nodes will receive more complete information in the case of limited sensor node processing capacity.

2 Simultaneous orthogonal matching pursuit(SOMP) algorithm

2.1 Distributed Source Coding [11]

According to the characteristics of the distributed sensor networks, a number of distributed coding algorithms are developed, these algorithms all involved the cooperation of sensor nodes, including predictive coding, distributed KLT and distributed wavelet transform, and also the three dimensional wavelet algorithm which uses inter-correlation between signal and in signal at the same time.

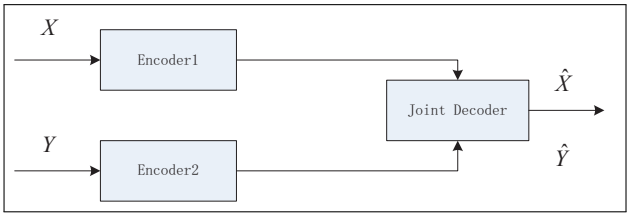


Figure 1: Distributed Source Coding.

Figure 1 is distributed source coding of two different source information. Processing of signal Y is completely independent, the transmission information is $H(Y)$, and processing of the

signal X need to use correlation with the signals Y. At the receiving side, the decoder $H(Y)$ can first restore the original signal Y; as the edge information Y, use the correlation between the information X and Y that contained in $H(Y)$, and send the joint information $H(X|Y)$, complete the joint decoding of the source information X, here with the sum of the amount of the transmission information X and Y must be less than the joint entropy $H(X|Y)$, so it can ensure the compression effect of source information. In one word, for distributed source coding theory, the key issue is how to use the original related characteristics of the signals X and Y, and then to do the coding and joint decoding independently.

2.2 SOMP algorithm and JSM-2 model

JSM-2 [3]: The signals in this model have different coefficient value, while they are all made up with the base vector of the same sparse coefficient set. This model can be applied some important applications, such as MIMO telecommunication and audio signal array. The signals which are accorded with JSM-2 model may be sparsity in the fourier domain, for example, degenerative frequency component which is caused by different propagation paths.

In JSM-2 model, the expression of the original signal shown by the following formula:

$$X_j = \Psi \cdot \Theta_j, j \in 1, 2, ..., J \tag{1}$$

Among them, the coefficient of each vector Θ_j, corresponding to a collection of indicators $I_j \subset 1, 2, ..., N$, and I_j only has K elements, namely, that is $\|I_j\|_{l_0} = K$, the sparsity of each original signal X_j is K.

Paper [7] presented the greedy algorithm to recover the related signal, which is called Simultaneous Orthogonal Matching pursuit (SOMP) algorithm. The algorithm is very similar to the OMP algorithm, but there are some small changes. SOMP algorithm is based on the concept of distributed source coding. The basic idea of SOMP algorithm is: assuming that the sensor nodes in WSNs are all consistent with the JSM-2 model, namely each part of the original signal includes only part of the information and do the sparse representation on the same sparse basis, but the sparsity of each of the original signal sparse are not the same.

Suppose that there are B distributed sampling signals $y_1, y_2, ..., y_B,$, change the second step of the OMP algorithm to find the index, using the index to solve the following simple optimization problem:

$$max_{\omega \in \Omega} \sum_{1}^{B} |\langle R_{k,t-1}, \varphi_j| \tag{2}$$

In which $R_{k,t-1}$ is the residual value of the first k distributed sample signal, the rest of the steps are the same with OMP algorithm. When $B = 1$, this process can be summarized as the standard OMP algorithm.

In conclusion, the SOMP algorithm is divided into the following steps:

1) Initialization: residual $r_0 = y$, the number of iterations $t = 1$, the index set $\bigwedge_0 = \phi$;

2) Find the index λ_t, solve a simple optimization problem;

3) Update the indexes set $\Lambda_t = \Lambda_{t-1} \cup \lambda_t$, and update the selected column space $\theta_t = [\theta_{t-1}, \varphi_{jt}]$. By convention, θ_t is an empty matrix;

4) By solving a least squares problem, get a new estimation of the signal;

5) Calculate the new approximate data and the new $a_t = \theta_t x_t$, $r_t = y - a_t$;

6) $t = t + 1$, if $t < m$, go back to step 2), otherwise, to step 7);

7) If $\|\theta_t\|_\infty > \gamma$, choose H_1, otherwise choose H_0. In which γ is nonzero threshold, $H_0 : \theta_s = 0$, $H_1 : \theta_s \neq 0$, θ_s is sparse coefficient.

3 DEEC clustering algorithm

DEEC [9] is a classical clustering algorithm, which can save the energy effectively and prolong the network lifetime. It can be able to construct optimized cluster structure for and data aggregation through the exchange of the message between the local nodes, and also balance network energy consumption effectively, then be better adapted to periodically collect data of sensor network applications. Simulation results show that DEEC can prolong the network lifetime of LEACH of about 45% under the smaller latency of network.

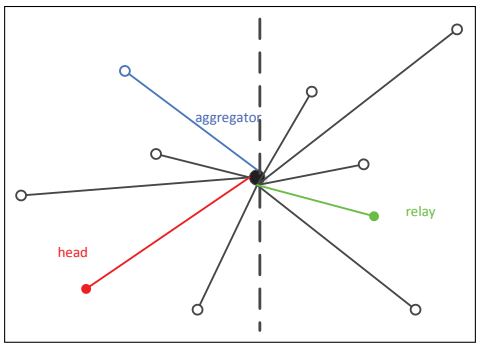


Figure 2: The structure of DEEC.

Clusters in the DEEC protocol contain four types of the nodes: the sensor, the head, the aggregator and the sender. Nodes in a network choose itself as the cluster head nodes with a certain probability value to form a cluster head node collection $Hhead_i$, become a cluster head node which is responsible for the set of clusters. The distance between each of the sensor nodes which are in the cluster i and the head nodes to meet the following conditions:

$$min_{1 \leq k \leq |H|} dist(v_{ij}, head_k) \tag{3}$$

But the head is not responsible for data aggregation, fusion and sending information to the base station immediately, it plays a role of a "call" in the first place. In the process of the set of clusters, we determine the aggregator and the sender through the calculation. The former is used to gather the information which is sent by clusters and then make the data processing and fusion, the latter accept the information transmitted by the former and then send them to the base station. Of course, set clusters, gather data and send information to the base station, those three types of task may be borne by one or two nodes.

The process of the head selection of DEEC is similar to LEACH: each node generate a random number, if it less than the threshold, the node will be the cluster head. But the determination of the threshold here is different from LEACH, DEEC neither consider the effect of the current round nor consider whether the node within several rounds of cluster heads, for DEEC, cluster head set H only have the effect of dividing the network into several clusters. The threshold of DEEC is the probability of each node to become cluster head P_{head}, $P_{head} = K_{opt}/N$, in which K_{opt} is the optimal value of number of clusters in the network; N is the number of nodes in the network during initialization. The derivation of K_{opt} has a related discussion in the literature [8], here only assume that it is a system parameter during the initialization time of the network. When a node is selected as the cluster head, it broadcasts messages $INVITE$ to

all other nodes, the nodes which receive the news choose the nearest cluster heads to $JOIN$ and send a message to its cluster head $JOIN(ID, Position)$. In this way, the entire network is divided into several clusters.

4 DCSH algorithm

In above algorithm, the ability of sensor nodes in WSNs is limited to only receive the information sent by a small amount of sensor nodes, and each sensor node has only its local information, therefore the joint decoding side cannot recover the full information, this will lead to the loss of the target source information when recover.

In this paper, based on our previous works [12] [13] [14], we propose an improvement scheme of DCS reconstruction algorithm (DCSH algorithm) which is based on SOMP. Its basic idea is: by using DEEC algorithm to divide sensor nodes into clusters, then select the cluster head, and the information of the nodes all gathered to the cluster head, then the information on the cluster head is the information within the entire cluster nodes. The sensor nodes send the information to the cluster head after data fusion and then use SOMP algorithm to recover the information. On one hand, the DCSH algorithm reduces the number of the nodes which transfer the information directly to the joint decoding end, avoids the loss of transmit information; on the other hand, by using the advantage of the DEEC protocol, our algorithm can achieve the goal of saving energy.

System model for DCSH algorithm is shown in Fig.3:

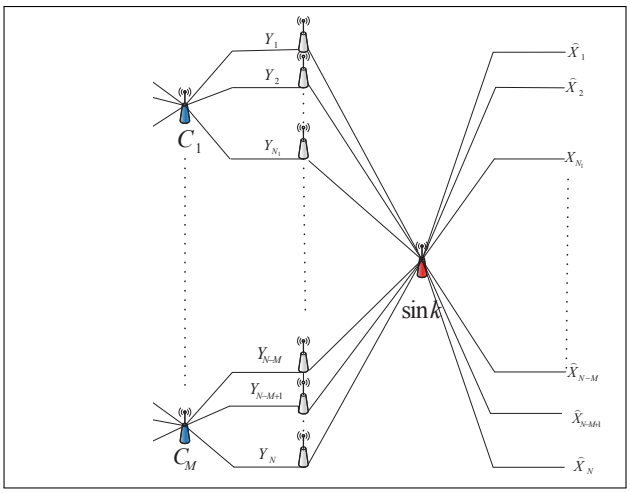


Figure 3: System Model for DCSH algorithm.

Suppose that there are N sensor nodes randomly distribute in the wireless sensor networks, meanwhile the sink node is in the center of the network area. There is a target source in the network, information on the target source will be sent to all nodes in the network, any sensor nodes can communicate with the sink node directly, and the signals all meet the joint sparse model JSM-2.

First, divide N nodes into M clusters, then elect the cluster heads $C_1, C_2, ..., C_M$ of M clusters respectively, and the number of the nodes in each cluster is $N_1, N_2, ..., N_M$, transmit the cluster information to each cluster head and then use DCS algorithm to encode information on the cluster head, getting the measured value $Y_1, Y_2, ..., Y_M$ which is transmitted to the sink node for the joint decoding. At last, get the reconstruction value $\hat{X_1}, \hat{X_2}, ..., \hat{X_N}$, thereby, recover the target source information accurately.

DCSH algorithm can be shown as follows:

1) Based on DEEC protocol, divide the sensor nodes into cluster, and choose the cluster head. The threshold P_{head}, $P_{head} = K_{opt}/N$ of DEEC is the probability value of each node to become cluster head, in which K_{opt} is the optimal value of number of clusters in the network; N is the number of nodes in the initialized network.

2) Determine the aggregator: when the head know location information of all nodes, aggregator can be determined through simple algorithm $FIND\_AGGREGATOR$ [10].

3) Determine the sender: using the algorithm $FIND\_SENDER$ [10].

4) Transmit the information of the sensor nodes to the cluster head, using their original measurement matrix to do the independently coding on the cluster head.

5) Use SOMP algorithms to joint reconstruction of the nodes information, thus recover the full information of the network.

5 The results of simulation and analysis

This paper uses Matlab as a tool for simulation, 100 sensor nodes are randomly distributed in deployment area , the center of the area is the cluster head, each of the node information conform to JSM-2 models.

First of all, we verify the performance of the SOMP algorithm. Get the reconstruction value $\hat{X}_1$ of the sensor node information X_1 by SOMP algorithm and OMP algorithm. The reconstruction error of is 5.1672×10^{-15} and 8.1596×10^{-15} . Figure 4 and Figure 5 show us the comparison between the two kinds of algorithms, and it apparently show that SOMP algorithm has the excellent effect of reconstruction. After compared with OMP algorithms, we will find SOMP algorithm has good performance and also the accuracy is higher.

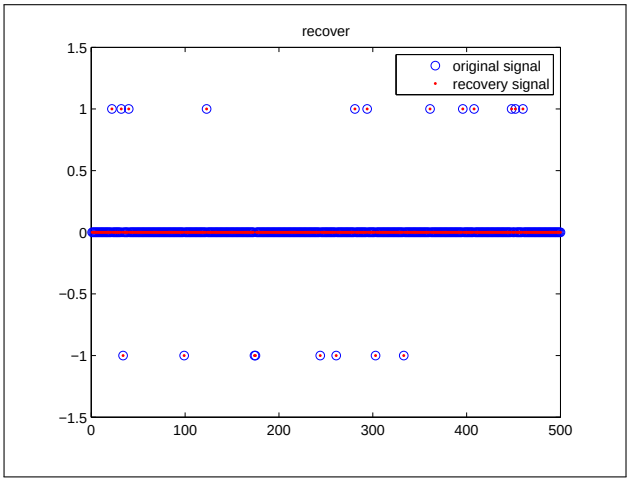


Figure 4: Comparison of SOMP algorithm.

Suppose that the LEACH protocol and DEEC protocol have the same number of the original data nodes, Figure 6 studies the performance under the conditions of same numbers of cluster nodes. In this figure, the two curves represent the protocol as LEACH and DEEC. Its specific results are shown in Figure 6, DEEC protocol has m less dead nodes When the algorithm run same time.

Known that the initial energy of the nodes is $50J$, the energy of the nodes decrease with the number of the election round, Figure 7 shows the relationship between the rounds and active

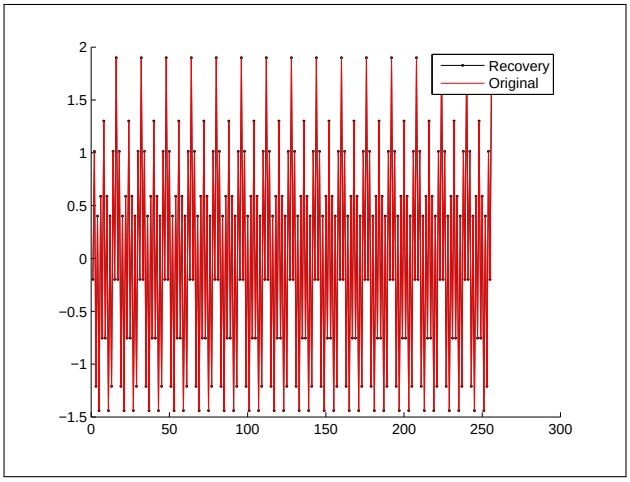


Figure 5: Comparison of OMP algorithm.

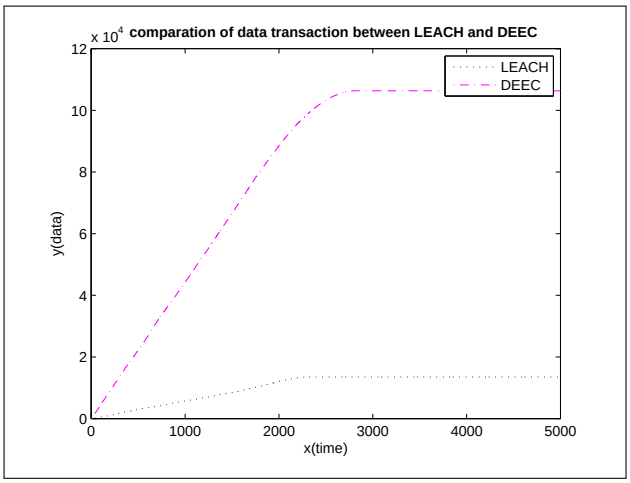


Figure 6: The comparison of the lifecycle.

nodes of DEEC algorithm. We can conclude that, in the same case, networks lifetime of DEEC is longer than traditional LEACH protocol, to a great extent, improve the network cycle.

For the further study of these two reconstruction algorithms, once again, we compare the average absolute error performance of the SOMP and OMP algorithm. Figure 8 demonstrates the relationship between the absolute error performances of the two algorithms, it can be seen that, in the case of the same number of measured values, the absolute error performance of SOMP algorithm is lower than the absolute error performance of OMP algorithm. To sum up, the comparison and analysis of the performance of SOMP algorithm and OMP algorithm in Figure 8 give a conclusion that in the same conditions, SOMP algorithm can ensure the accuracy rating of the reconstruction while reducing the number of measured values. In this way we can save the energy of the network and improve the efficiency of the signal reconstruction.

6 Conclusions

This paper proposes a distributed compressed sensing algorithm for hierarchical Wireless Sensor Networks (DCSH algorithm). This algorithm utilize DEEC algorithm to get the accurate information by recovering the information of sensor nodes in the network. The sensor nodes

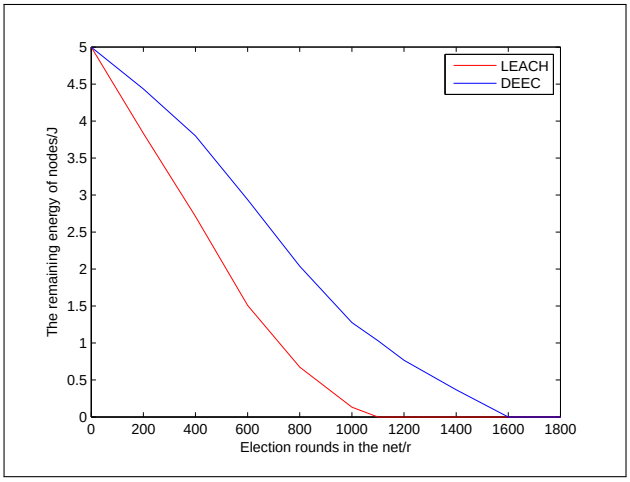


Figure 7: The comparison of the nodes and the wheel.

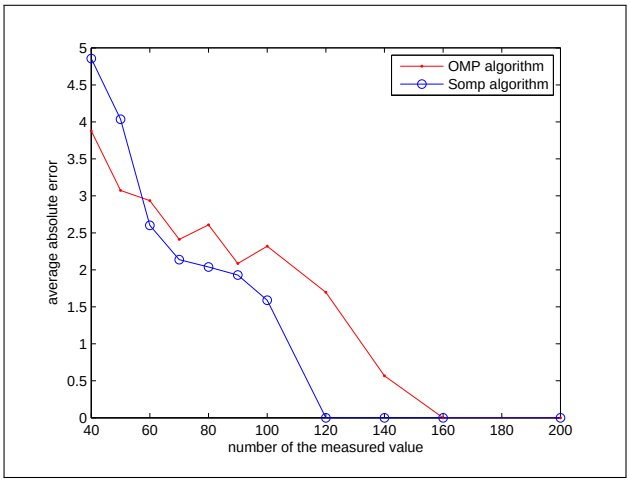


Figure 8: The change of two kinds of algorithm.

choose the cluster head by DEEC protocol. We take the residual energy of nodes into consideration while in the process of choosing the cluster head, prevent some nodes energy exhausted prematurely, this can not only extend the network lifetime effectively, but also reduce the number of the nodes which transmit data directly to the gathering node. Then based on the spatial correlation of the cluster nodes and joint sparse model (JSM-2), by using distributed compressed sensing reconstruction algorithm(SOMP), recover the original node information from a small amount of information accurately. The simulation results show that this algorithm has better performance, which can improve the network lifetime, and can be better adapted to periodically collect data of the WSNs application.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant No. 61063037 and No. 51364001, Open Program of State Key Laboratory Breeding Base of Nuclear Resources and Environment under Grant No.NRE1102, Key Projects in the Science and Technology Pillar Program of Jiangxi Province of China under Grant No. 20111BBG70031-2 and 20133BBE50033, Educational Commission of Jiangxi Province of China under Grant No.

GJJ13335 and GJJ13354, and Foundation for Young Scientists of Jiangxi Province of China under Grant No. 20133BCB23016.

Bibliography

[1] Donoho D L.(2006); Compressed sensing, *IEEE Transactions on Information Theory*, 2006, ISSN 0018-9448, 52(4): 1289-1306.

[2] Candés E.(2006); Compressive sampling, *In: Proceedings of International Congress of Mathematicians*, ISBN 978-3-03719-022-7, 1433-1452.

[3] D.Baron et al (2005); Distributed compressive sensing, *Technical Report* ,pre-print.

[4] D.Baron et al (2005); An Information Theoretic Approach to Distributed Compressed Sensing, *In: Conference on Communication, Control, and Computing* , ISBN: 9781604234916.

[5] M.F.Duarte et al (2005); Distributed Compressed Sensing of Jointly Sparse Signals, *In: Proceeding of the 39th Asilomar Conference on Signals, Systems and Computation* , ISSN 1058-6393, 1537-1541.

[6] J Tropp; A. Gilbert; M Strauss(2006); Algorithms for simultaneous sparse approximation, Part I: Greedy pursuit, *Journal of Signal Processing*, ISSN 0165-1684, 86: 572-588.

[7] W Dai; O Milenkovic(2009); Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, ISSN 0018-9448, 55(5): 2230-2249.

[8] L.Zhao; L.Q.Lian(2005); Distributed and Energy Efficient Self-organization for On-off Wireless Sensor Networks, *International Journal of Wireless Information Networks*, ISSN 1068-9605, 12(1) : 211-215.

[9] L.Qing; Q.Zhu; M.Wang(2006); Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks, *ELSEVIER, Computer Communications*,ISSN 0140-3664, 29(12): 2230-2237.

[10] Y.X.Liu et al(2010); Regularized Adaptive Matching Pursuit Algorithm for Signal Reconstruction Based on Compressive Sensing, *Journal of electronics and information*, ISSN 1009-5896, 32(11): 2713-2717.

[11] D. Slepain; J. K. Wolf(1973); Noiseless coding of correlated information sources. *IEEE Transaction on Information Theory*, ISSN 0018-9448, 19(9): 471-480.

[12] Nan Jiang(2014). WDEM: Weighted Dynamics and Evolution Models for Energy-Constrained Wireless Sensor Networks, *Physica A: Statistical Mechanics and its Applications*, ISSN 0378-4371, 404: 323-331.

[13] Nan Jiang; Sixin Jin; Yan Guo; Yueshun He(2013); Localization of Wireless Sensor Network Based on Genetic Algorithm, *International Journal of Computers Communications & Control*, ISSN 1841-9844, 8(6): 825-837.

[14] Nan Jiang; Rigui Zhou; Qiulin Ding(2009); Dynamics of Wireless Sensor Networks, *International Journal of Distributed Sensor Networks*, ISSN 1550-1329, 5(6): 693-707.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):439-452, August, 2014.

Performance Evaluation of ENUM Name Servers

S.H. da Mata, J.M.H. Magalhaes, A. Cardoso, H.A. Carvalho, P.R. Guardieiro

Saulo Henrique da Mata*, Johann M. H. Magalhaes,
Alexandre Cardoso, Paulo R. Guardieiro
Federal University of Uberlandia
Av. Joao Naves de Avila, 2121, Bloc 3N
Uberlandia, MG, Brazil - 38400-912
johann@iftm.edu.br, alexandre@ufu.br, prguardieiro@ufu.br
*Corresponding author: saulodamata@doutorado.ufu.br

Helio A. Carvalho
CPqD, Rua Dr. Ricardo Benetton Martins, s/n
Parque II do Polo de Alta Tecnologia, Campinas, SP, Brazil - 13086-902
helioac@cpqd.com.br

Abstract: ENUM is a protocol designed for mapping E.164 telephone numbers into Internet URIs. This protocol imposes new requirements and challenges on traditional DNS servers. In this sense, this paper presents a performance evaluation of ENUM name servers considering these new scenarios. We selected four name servers implementations (BIND, MyDNS-NG, NSD and PowerDNS) to compare their performance, using a benchmarking testbed. To perform the evaluation, we have defined three performance metrics: query throughput, response time, and CPU usage. These metrics required the development of a new procedure for the evaluation, since DNSPerf presented limitations in the management of CPU resources. From results, we classified MyDNS-NG as not suitable for ENUM purposes, due to its quite low query throughput and poor scalability. PowerDNS server performed better than MyDNS-NG, but presented high sensibility to database size. On the other hand, BIND and NSD servers achieved a high query throughput with great scalability features. Since we have used extended scenarios and size of database that were not evaluated before, these results bring to light an interesting insight about ENUM servers.
Keywords: ENUM, Benchmarking, BIND, NSD, MyDNS-NG, PowerDNS

1 Introduction

Nowadays, we are witnessing a continuous evolution of communication networks. In this evolution process, services previously exclusive to one communication network began to be shared with other networks. In this context, Voice over IP (VoIP) is one of the best examples of this migration, where the voice service from Public Switched Telephone Network (PSTN) is added to data service from Internet and both services are available with a lower cost than the currently one offered by traditional networks.

Despite the expanding number of VoIP users and the availability of several providers, geographically, users of this technology are located on a set of "isolated islands", where the interconnection among these islands may occurs by means of PSTN. This connection is performed by Media Gateways (MGWs) and E1 links, which are commonly used to interconnect the VoIP carriers with the PSTN. This interconnection infrastructure requires investments by operators in equipment and the payment of fees for the use of communication links. These expenses increase the total operating costs, burdening the end user.

Furthermore, one should consider that in spite of the advantages of VoIP application over the PSTN, it is not possible the immediate replacement of this technology, because of the complexity involved. This implies a transition period when both technologies will coexist, since the migration

of PSTN subscribers to VoIP can occur slowly and gradually. In this scenario, we have four basic types of calls: PSTN-PSTN, PSTN-IP, IP-PSTN and IP-IP. So, the problem is not restricted only to avoid the use of PSTN for IP-IP calls, but also to allow calls between terminals of two different networks (PSTN-IP, IP-PSTN).

Traditionally, E.164 telephone numbers [1] are used in PSTN and Uniform Resource Identifiers (URIs) [2] are used on the Internet. Signaling protocols such as SIP and H.323 can work with these two types of identifiers. For Internet users, who usually have the availability of a alphanumeric keyboard, the textual names (URIs) are mostly used because of easy memorization and deduction. However, a problem arises when VoIP and PSTN networks are interconnected, since PSTN terminals typically provide a numeric keyboard to users. Thus, it is necessary a mapping system between E.164 telephone numbers and URIs which permits the interconnection of PSTN and VoIP users. Moreover, this mapping system must allow IP-IP calls without the connection with PSTN. In this context, the use of Electronic Number Mapping (ENUM) protocol stands out as a plausible solution.

ENUM [3] is a technology that maps E.164 telephone numbers used in PSTN into URIs used on the Internet. Although ENUM is based on existing DNS infrastructure, it has certain features, for example, a database with a huge number of records and the requirement of short resolution time to reach a PSTN similar performance. In addition, it causes an increase in network traffic, since the size of the response message to ENUM queries is greater than in classic DNS. Therefore, it is of fundamental importance a study of how ENUM servers perform on this new environment.

In this context, this paper presents a new performance evaluation of ENUM name servers. Using a high performance hardware, we selected four different name servers implementations (BIND, MyDNS-NG, NSD and PowerDNS) for the evaluation. Our goal was to make a detailed study of the performance of each one of these servers and verify whether this performance can meet the ENUM protocol requirements.

The rest of the paper is organized as it follows: Section 2 describes related work. In Section 3 we present our testbed and methodology. Section 4 shows the benchmarking results. Finally, concluding remarks are offered in Section 5.

2 Related work

The performance study of ENUM servers has been reported in some works. In [4] and [5] we find a benchmarking study of three ENUM servers (BIND, Navitas and PowerDNS). According to authors, Navitas and PowerDNS met the requirements of ENUM protocol, considering performance metrics as query throughput, response time and scalability. On the other hand, BIND was not qualified as a serious ENUM server.

In [6], we find another benchmarking study of two ENUM servers (BIND and NSD). In this work, BIND and NSD presented a similar performance for query throughput that is quite different from the results presented in [5].

In [7], there is an evaluation study of number portability and record updating process in ENUM systems. Using DNS zone transfer method for database synchronization, this work evaluated two ENUM servers (BIND and PowerDNS). The considered metric was the information transference speed between different zones. The results showed that BIND was 3 to 4 times faster than PowerDNS to deliver the number portability data.

In [8] there is a performance evaluation of ENUM protocol considering different approaches for ENUM architecture. In this work, the focus was not the ENUM server performance, but the hierarchy arrangement of these servers.

In [9], there is not a comparative study between ENUM servers, but BIND was evaluated in the context of French ENUM architecture. The goal of this work was to build a simulation tool

based on the data of a real ENUM system. A similar work can be found in [10].

In all these works, we found different types of hardware and servers. Besides, there are differences in the methodology and, as a consequence, it is not easy to compare the results presented by each work. However, among these works, [5] and [6] present a more complete study, a concise methodology, and good hardware and software profiles. Although these works present similar objectives, they report divergent results. Moreover, they were developed about 4 to 5 years ago. Since then, we have been observing a great evolution for hardware, software and database technologies, creating the demand for updated and reliable studies about the performance of ENUM servers.

In [11], one can find a preliminary evaluation of ENUM name servers. In the present paper, we extend this work by adding new scenarios of evaluation not contemplated in any previous works, for example, performance evaluation considering multiples Naming Authority Pointer (NAPTR) records per Fully Qualified Domain Name (FQDN).

In this sense, our study fulfills and updates all these former works. We built a powerful testbed that outperforms the hardware and software used in [5] and [6]. We improved the methodology presented in [5] by extending the range of the parameters of evaluation. Besides, we report limitations of the DNSPerf (version 1.0.0.1) and present results for new scenarios or evaluation.

3 Testbed and methodology description

3.1 Performance Metrics

Three performance metrics were defined for the tests:

- Query throughput: the number of queries which the ENUM server can respond successfully per unit of time, i.e., queries per second (qps).
- Response time: time between sending of query to server and the moment the client receives the answer to that query.
- CPU and RAM usage: this metric evaluates how well the ENUM server uses these hardware resources.

3.2 ENUM servers

Nowadays, there is a great number of existing ENUM servers. We selected a set of servers in a way that we could evaluate different approaches for implementation of a name server. We also considered other factors such as the relevance of the server, availability, documentation and updating process. In this sense, we selected BIND (version 9.8.1) and NSD (version 3.2.10) to represent the name servers based on zone files. MyDNS-NG (version 1.2.8.31) and PowerDNS (version 2.29) were selected to represent ENUM servers that allow the use of other types of database. For these servers we used MySQL (version 5.1) as the database.

3.3 EnumBenchTool

EnumBenchTool is a test management tool for benchmarking of ENUM servers developed by the Network Computer Lab (NCL) of the Federal University of Uberlandia. This tool has been developed to automate, standardize and validate the tests and to facilitate the achievement of results.

At the current stage of development, EnumBenchTool admits the servers BIND, MyDNS-NG, NSD and PowerDNS. This tool is not responsible for the benchmarking test itself, but it packs several other existing tools, simplifies the benchmarking management and makes configuration, synchronization and validation processes transparent to user. Among these packed tools, we highlight DNSPerf [12], which is a software developed by Nominum, widely used to analyze the performance of authoritative DNS servers. EnumBenchTool is responsible for triggering each DNSPerf test step, setting the number of emulated clients, the runtime, the query file and other parameters. In addition, EnumBenchTool is responsible for processing the results from DNSPerf.

3.4 Number of DNSPerf processes

Unfortunately, DNSPerf (version 1.0.0.1) is not a multi-core software, i.e., it uses only one core of the processor for each operating system (OS) process. Depending on the performance of the ENUM server, DNSPerf may not be able to send a sufficient amount of queries to saturate the name server using one single OS process. Thus, the saturation occurs on the client, invalidating the obtained results. So, it is necessary to increase the number of DNSPerf OS processes emulating clients to avoid this saturation. However, this increase cannot be indiscriminate, since the operating system has limitations to manage a very large number of processes. Initial tests showed that the use of five processes is a reasonable choice for this type of evaluation, as showed in Table 1.

Table 1: Impact of the number of DNSPerf processes in server throughput.

Processes	Clients/Process	Max. Throughput (qps)
1	500	202743
5	100	309927
10	50	305523
20	25	95189
500	1	60559

3.5 Number of clients

In [5], there is an evaluation interval from 2 to 60 clients. However, initial tests proved that this interval is too short to evaluate the server satisfactory, since 60 clients were not enough to saturate the ENUM server and, then, the results would not be plausible. Therefore, we established that the number of emulated clients, who make simultaneously queries to server, would be in a range of 2 to 1000 clients. This range proved to be sufficient for stability of the system.

3.6 Type of records

We evaluated the ENUM servers considering two types of records:

- existing records: the name server has a DNS NAPTR record for the FQDN queried;
- non-existing records: the name server has not a DNS NAPTR record for the FQDN queried.

3.7 Database record sets

To check whether the server meets the requirement for scalability, three database record sets with 500 thousand (500k), 5 million (5M) and 50 million (50M) of records were defined.

3.8 Query files

Using the EnumBenchTool, we created query files for each database record set and for each record type as well. For each query file, these FQDNs were generated in a random manner to avoid cache memory interferences. In this sense, we also generated independent random query files for each DNSPerf processes.

3.9 DNS zones

Each record set was divided into ten zones, with each zone consisting of one tenth of the total number of records. This division method, reported in [5], is a practical and efficient generation process for different types of records. Furthermore, using BIND, we evaluated the impact of the number of DNS zones in the server query throughput, as showed in Table 2.

Table 2: Impact of the number of DNS zones in server throughput.

Number of DNS Zones	Max. Throughput (qps)
5	280994
10	324550
50	271256
100	270402
1000	265486

Each zone file starts with a $TTL directive. We set the TTL as 86400 seconds. As we have a huge database and the records are queried using random and independent query files, the cache mechanism is not a great factor in our experiment. An example of the beginning of a zone file is as follow:

```
\$TTL 86400
\$ORIGIN 0.3.3.4.3.5.5.e164.arpa.
@ IN SOA ns.0.3.3.4.3.5.5.e164.arpa. root.e164.arpa. (
                                                      2011080901
                                                      21600
                                                      3600
                                                      604800
                                                      3600
                                                     )
 IN NS ns.0.3.3.4.3.5.5.e164.arpa.
ns IN A 10.0.0.1
```

3.10 Testbed layout

The testbed consists of two Linux machines, running Ubuntu 11.10 and Ubuntu Server 11.10. Each machine has two processors Intel Xeon 6-Core E5645 HT 2.40 GHz, 16 GB of RAM and two hard drives of 500 GB. Both machines were connected via 1000 Mbps Ethernet connections. The testbed layout is based on [5] and is illustrated in Figure 1.

The master entity is responsible for synchronizing the activities of the machines under test. It sends commands to other two machines for the starting of each step of the test as well as the resource monitoring. Control messages were sent through a link of 100 Mbps. Queries and responses were sent through an exclusive link of 1000 Mbps.

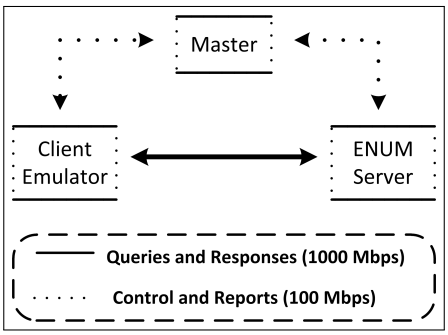


Figure 1: Testbed layout.

4 Performance comparison of the ENUM servers

The tests carried out at NCL were constituted of multiples steps. For each of these steps, the number of emulated clients was gradually incremented, according to commands sent by the master entity. Each step had a duration of 60 seconds, in which DNSPerf sent queries to ENUM server. At the end of the step, there was a timeout of 10 seconds before the starting of the next step to assure that the queries from the earlier step would not interfere in the results of the current step.

4.1 Overall performance

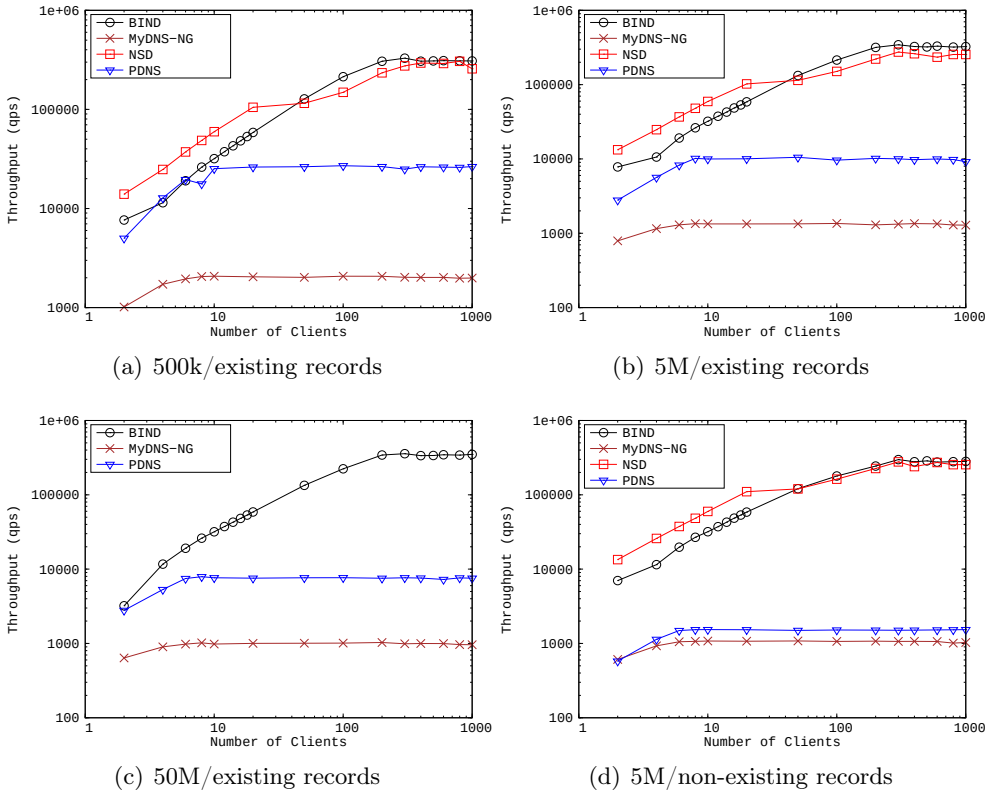
Figure 2 shows ENUM servers performance for query throughput considering the record sets and record types previously defined. In Figure 2(a), it is possible to observe that BIND and NSD presented a great performance for throughput, reaching a maximum throughput of about 300 kqps. In this same figure, we can observe that PowerDNS and MyDNS-NG presented a maximum throughput much lower than the servers based on zone files. This occurred because servers that use zone files load all the records to physical memory (RAM), which is known by its access speed. In our tests, MySQL was not able to follow this speed.

Figures 2(b) and 2(c) show the performance of ENUM servers for throughput with an increased database. BIND kept its performance even with a huge increase of the database. NSD also presented a great performance with 5M of records in the database, but the compiler of NSD zone files (*zonec*) failed to build all zones for the 50M record set. This failure of *zonec* was also reported in [6]. PowerDNS and MyDNS-NG were sensitive to database increasing and the throughput decreased as we increased the database size.

Another important parameter to be analysed is the server response time. To offer a service with quality similar as PSTN, VoIP call establishment time should not be longer than the PSTN call signalling interval. ENUM lookup response time is part of the total signalling interval for VoIP call setup. According to [13], the PSTN signalling interval has mean of 205-218 ms (337-350 ms in the 95th percentile). Thus, the lookup response time on the ENUM server should be lower than this mean value, since successive queries may occur in a single call setup.

Figures 3(a), 3(b) and 3(c) show the average response time for the evaluated ENUM servers. We note that all evaluated servers were able to satisfy the requirement of lookup response time in a VoIP call. This figure is also in consonance with the throughput results earlier commented. Servers that use zone files are faster than those that use MySQL database and, as consequence, the former servers presented a higher value for query throughput.

Figures 2(d) and 3(d) show a similar scenario, but with queries to non-existing records. In this scenario, BIND and NSD kept their performance. On the other hand, PowerDNS and



(a) 500k/existing records (b) 5M/existing records

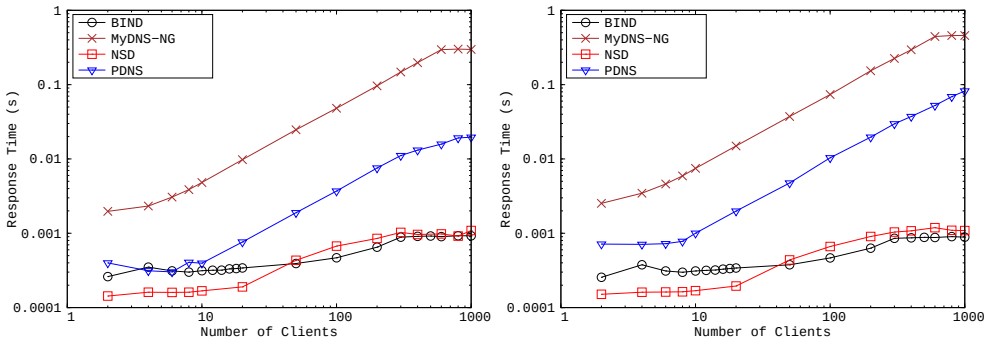
(c) 50M/existing records (d) 5M/non-existing records

Figure 2: Query Throughput.

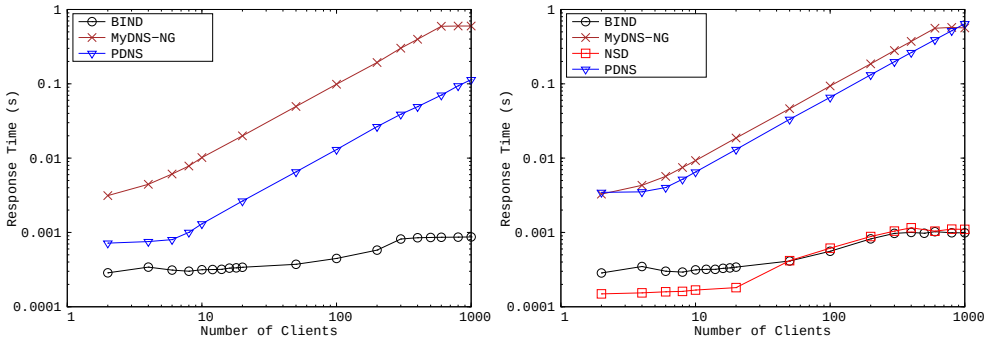
MyDNS-NG presented a reasonable decrease in their performance when we considered queries to non-existing records.

Figures 2 and 3 indicate that BIND and NSD reached saturation when 300 clients were sending queries to ENUM server. Under the saturation condition, the throughput of the server stops to increase and enters an unstable zone. Similarly, the response time stops increasing under this condition. In this situation, these servers start to experience increasing losses, as showed in Figure 4. However, even under saturation condition these servers answers the queries with a response time near those found for maximum throughput and, therefore, the response time under saturation condition shows only a small variation. On the other hand, MyDNS-NG reached saturation before 10 active clients in the system. This throughput low performance is reflected in response time that continues to increase until the system starts to experience losses. PowerDNS also reached saturation before 10 active clients in the system, but it presented a different behavior for response time and loss rate. The response time continued growing up even when losses began to be perceived.

Figure 5 indicates servers CPU usage. We can note that none of the evaluated servers have overpassed the maximum CPU capacity of the server, confirming that server saturation was caused by the software and not by hardware restrictions. In Figures 5(a) to 5(d), we can also note that to reach a high query throughput, BIND uses up to 90% of server CPU resources. Surprisingly, NSD reached a similar throughput with approximately 15% of CPU resources. The results for MyDNS-NG and PowerDNS indicate that these servers did not take advantage of the capacity of multi-core processors. We believe that this is the main factor for the differences in performance between the ENUM servers. While BIND and NSD used all the cores of the processors, PowerDNS and MyDNS-NG used only one of the cores. The implementation details



(a) 500k/existing records (b) 5M/existing records



(c) 50M/existing records (d) 5M/non-existing records

Figure 3: Response Time.

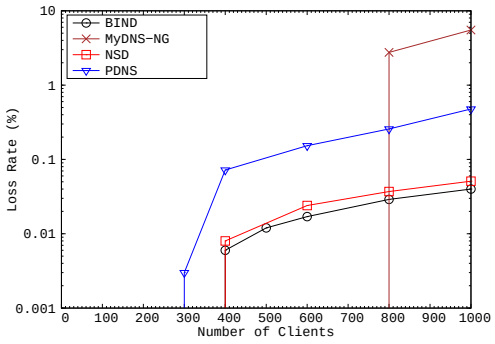


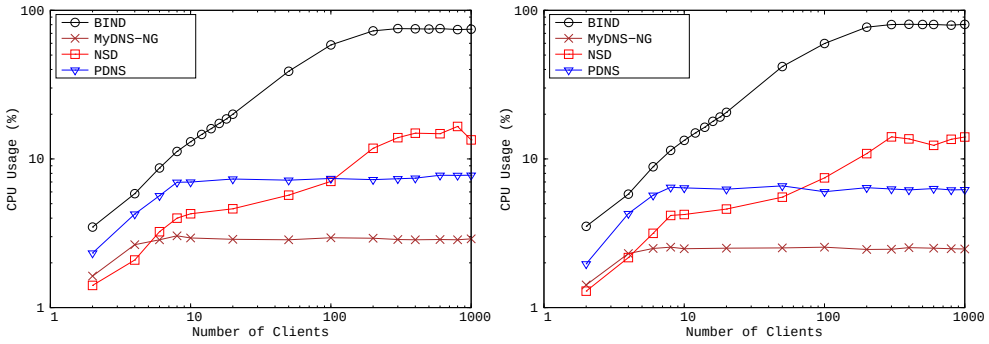
Figure 4: Loss rate considering 5M/existing records.

that cause these different levels of CPU and memory resource utilization are out of the scope of this work.

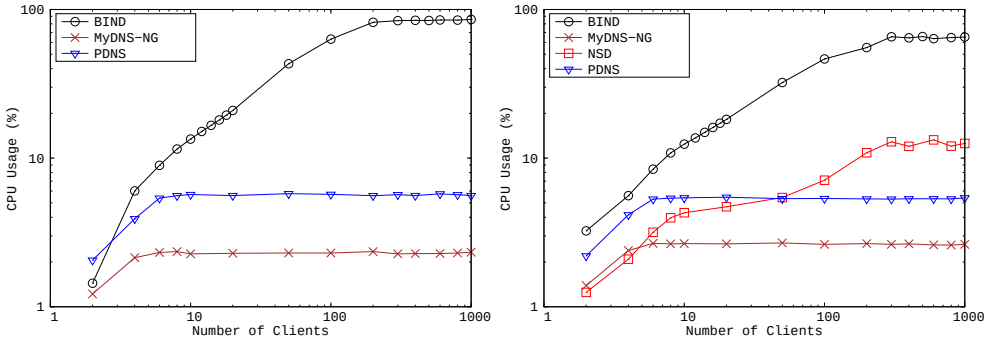
We also observe that there is only a slight variation in CPU usage when we evaluate different size of database and different types of records as well.

Table 3 shows memory usage for each ENUM server. BIND and NSD are based on zone files and, as a consequence, the greater the size of database, the greater the consume of RAM. On the other hand, MyDNS-NG and PowerDNS rely on MySQL as their backend. Thus, the RAM usage is smaller than BIND and NSD, with small variation when the database is increased.

In order to verify if other variables interfered on the results, two auxiliary metrics were considered: DNSPerf CPU usage and network usage. Figure 6(a) shows that only part of the total CPU resources were used by DNSPerf on the client machine, excluding the possibility of



(a) 500k/existing records (b) 5M/existing records



(c) 50M/existing records (d) 5M/non-existing records

Figure 5: CPU usage.

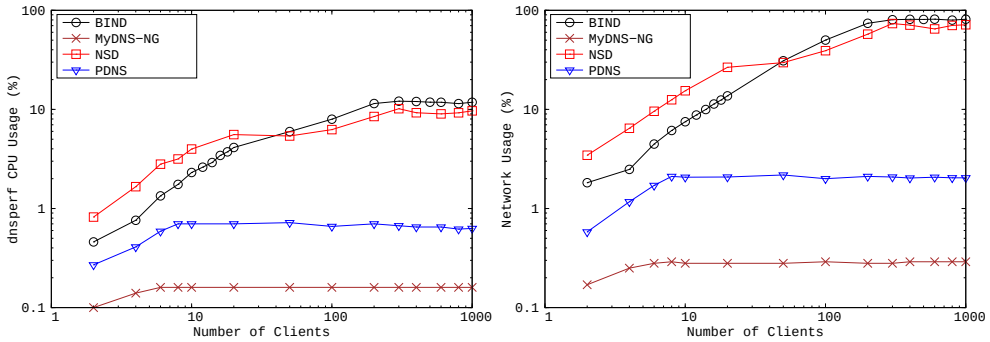
Table 3: Memory Usage.

Records	BIND	MyDNS-NG	NSD	PowerDNS
500k	1.1%	6.9%	1.2%	8.6%
5M	8.6%	6.9%	11.5%	8.9%
50M	82.8%	6.9%	Failed	9.8%

saturation of hardware resources on the client. Similarly, Figure 6(b) illustrates the network usage, and we can observe that the average value of usage has not overpassed the maximum capacity of the link (1000 Mbps).

4.2 Performance for different types of database

Until now, BIND and NSD has overcome MyDNS-NG and PowerDNS in all scenarios. These servers represent two different approaches of database for name servers, namely, zone files and MySQL. In this context, we built a new scenario to evaluate the effect of these different approaches in the performance of the server. To accomplish this task, we chose PowerDNS that is the most flexible server among the evaluated servers. PowerDNS can work with a great number of different backends. Therefore, we evaluated this server with two different databases: zone files and MySQL. Figure 7 shows the evaluation results. PowerDNS had a better performance when we used zone files. However, even with this improvement, PowerDNS performance is quite lower than BIND performance. This occurs mainly because PowerDNS has a poor CPU usage.



(a) DNSPerf CPU usage (b) Network usage

Figure 6: DNSPerf CPU and network usage considering 5M/existing records.

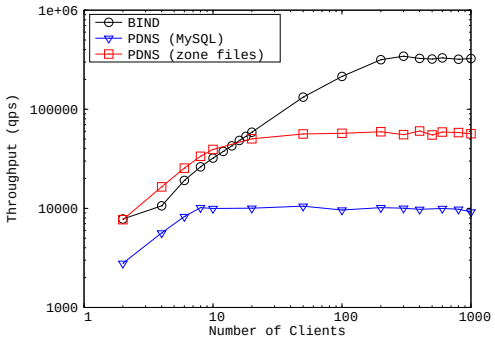


Figure 7: Query throughput performance for PowerDNS (MySQL and zone file backends) and BIND (zone files backend).

4.3 Performance considering the use of swap memory

Despite the great performance of servers based on zone files, they have some drawbacks. Each time the name server is restarted, it must load all records to RAM. Depending on the size of the record set, this can take a long time. Moreover, in ENUM context, the update process of a record stored in zone files is not straightforward, even when we use DNS Dynamic Updates [14]. Other relevant factor is the available RAM of the server. Servers based on zone files rely on RAM to reach high performance. If RAM is not enough to load all the records stored in zone files, the OS will use swap memory to load the rest of data. However, swap memory is quite slow when compared to RAM speed. Therefore, if swap memory is used, the performance of the ENUM server will decrease. In order to study this phenomenon, we evaluated BIND in a scenario where the RAM is not enough to load all records in the zone files. For this scenario we used 100M of existing records. Figure 8 shows the result. As we have foreseen, the query throughput is severely affected. The throughput using swap memory is about 300 times lower than in the scenario using only RAM.

4.4 Performance considering multiples NAPTR records per FQDN.

ENUM protocol is responsible for mapping E.164 telephone numbers into URIs. In this context, the E.164 telephone number is converted into a FQDN for querying process. However, one FQDN can be mapped into multiples NAPTR records, according to available services for a given user. Thus, it is interesting to know the behavior of ENUM servers in a scenario where

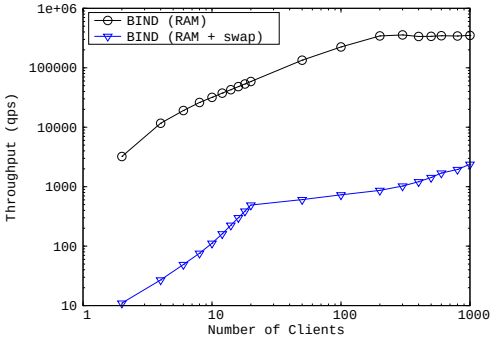


Figure 8: BIND using only RAM and RAM plus swap memory.

there are multiples NAPTR records for each FQDN.

In our evaluation, the number of NAPTR records per FQDN can be different for each server, since each of these servers presented distinct sizes of response messages. The number of NAPTR records in each response message was incremented until the size of response message was near to 512 bytes, which is the typical maximum value of a UDP segment. If a message is bigger than 512 bytes, this message is truncated. In this scenario we used only existing records and the 5M record set.

Figure 9 shows the evaluation results. In Figure 9(a), BIND presents the same performance for both evaluated cases until 100 clients. From this point, for 7 NAPTR/FQDN curve, the throughput remains constant, indicating a possible saturation state. However, analyzing Figure 10(a), it is possible to see that the saturation occurred on the communication link and not on the server. In this case, it would be necessary a link with more capacity to get reliable results. On the other hand, this experiment shows that BIND has a great performance even with more NAPTR records to manage. Figures 9(b) to 9(d) show that the other servers were affected by the increase of response message size. Differently from BIND case, for these servers the communication link was not saturated (see Figures 10(b) to 10(d)).

4.5 Comparison with related work

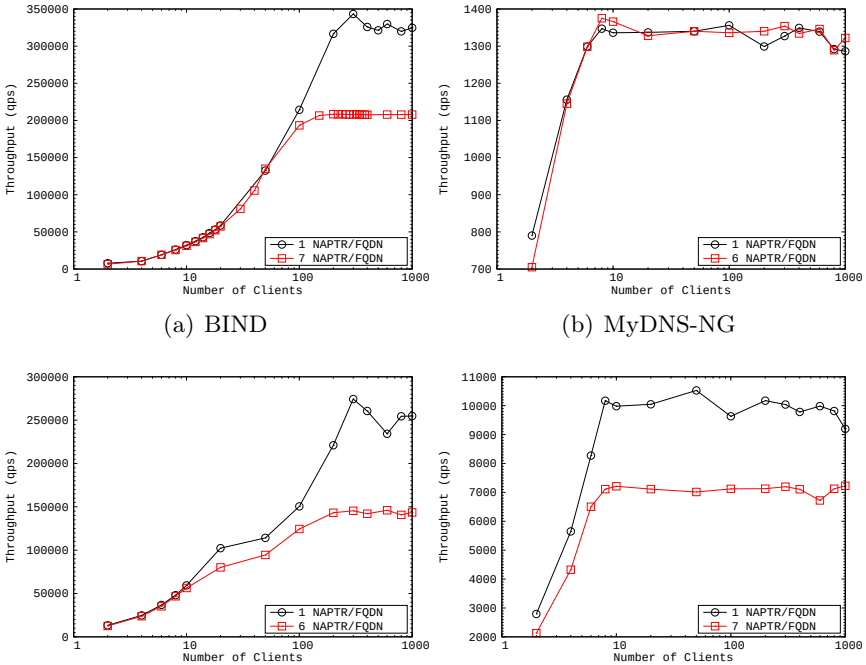
The comparison between different research works is not an easy task, because of different methodologies or lack of details about the experiment. Nevertheless, [5] and [6] are similar works that allowed us to perform some comparison to verify the relevance of our results.

Table 4 shows the comparison of the approximated maximum throughput for the three researches. In this comparison, our results and results from [5] are related to a 5M record set. For [6], the results correspond to a 1M record set. For all cases we used existing records.

Table 4: Comparison of the maximum throughput.

	Maximum Throughput (kqps)		
Work	BIND	NSD	PowerDNS
this	325	275	10
[5]	0.650	not evaluated	5.5
[6]	50	40	not evaluated

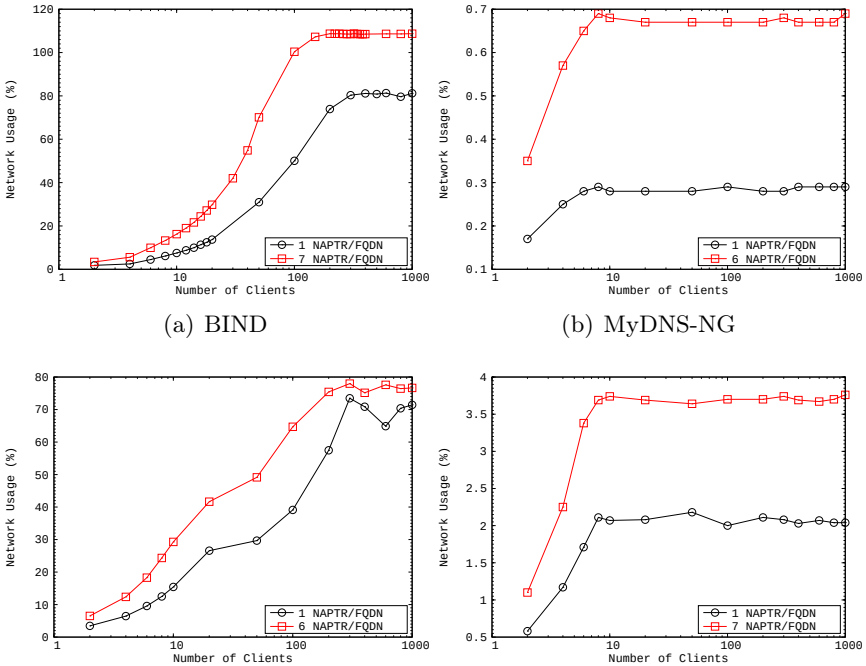
From Table 4 we observe the huge increase of throughput that we achieved in our experiments. Results are specially surprisingly for BIND due to its great evolution. Authors in [5] evaluated BIND version 9.2.4 and classified it as not suitable for ENUM due to its poor scaling properties.



(a) BIND (b) MyDNS-NG

(c) NSD (d) 800 items

Figure 9: Query throughput performance considering multiples NAPTR records per FQDN.



(a) BIND (b) MyDNS-NG

(c) NSD (d) PowerDNS

Figure 10: Network usage considering multiples NAPTR records per FQDN.

In [6], the test was performed with BIND version 9.4.0 and the throughput was more than 75 times greater than in [5]. In our evaluation, BIND version 9.8.1 presented the best scaling

properties among the evaluated servers and the highest throughput that was 500 times greater than the throughput reported in [5]. NSD also presented a significant evolution from version 3.0.4 to version 3.2.10. It is important to mention that this increase in throughput is also caused by the evolution of the hardware, since we used a more powerful testbed than the former works. These results brings good expectations for the future of ENUM protocol, since the performance of the ENUM servers are getting improved, allowing the deployment of new services.

5 Conclusions

ENUM protocol has been considered as the most promising solution for the integration between the different existing communication networks. In this sense, in this paper, we have selected four different ENUM servers, namely, BIND, MyDNS-NG, NSD and PowerDNS to study their behavior in the new scenarios of deployment. We chose BIND and NSD to represent zone files name servers. MyDNS-NG and PowerDNS were selected to represent name servers with SQL database.

The evaluation was performed in a testbed with high performance machines, and the last stable versions of the ENUM servers selected for the benchmarking. We also developed a new tool to manage our tests. The EnumBenchTool proved to be of fundamental importance to facilitate and validate our work.

In our evaluation scenarios, we extended the range of clients sending queries to the server in a far greater range than the ones used in previous works.

From results, we identified that MyDNS-NG presented the worst performance in all scenarios. We believe that this server is not mature enough to be used in an ENUM context.

PowerDNS performed better than MyDNS-NG. However, due to its inefficient use of CPU and RAM memory resources, it performed quite lower than BIND and NSD.

BIND presented the highest query throughput and a great scalability features. This is a totally new and different result when compared with results from [5]. Our results proved the evolution of this server in the last years.

NSD had the most efficient CPU usage performance. It presented a query throughput similar to BIND, but with a lower memory consume. On the other hand, NSD presented problems to store a huge amount of records, compromising its scalability.

We also evaluated the impact of the response message size in the servers performance. This is a new scenario that was not contemplated in any of previous works. BIND proved its superiority with a better performance when compared with the other ENUM servers.

Another important scenario is the evaluation of the performance of ENUM servers, that use zone files, when the database is greater than the RAM capacity. It is worth to say that zone files allow the server to reach a higher throughput than MySQL only if the name server has enough RAM to load the records stored in the zone files. Moreover, the update process is not straightforward when we use zone files.

Beyond the new results presented in this paper, we also highlight the improvements in the methodology for benchmarking of ENUM servers. We have identified that DNSPerf saturates when only one OS process is used to query the ENUM server. This limitation of DNSPerf brings to light the necessity of improvement in this tool to follow the hardware advances. To the best of our acknowledge, we are the first ones to report this limitation.

In short, we believe that we got important results for the benchmarking of ENUM servers research field. We presented new information about the performance of important ENUM servers and clarified some divergences between former works, as in [5] and [6]. Moreover, we improved the methodology for the benchmarking of ENUM servers and presented results for scenarios that were not contemplated in any other works.

Acknowledgement

The authors gratefully acknowledge financial support from the Studies and Projects Finance Organization (FINEP), under project No. 01.10.0560.00.

Bibliography

[1] ITU-T (2005). The International Public Telecommunication Numbering Plan. ITU Recommendation E.164.

[2] Berners-Lee, T.; Fielding, R.; Masinter, L. (1998); Uniform Resource Identifiers (URI): Generic Syntax. RFC 2396.

[3] Bradner, S.; Conroy, L.; Fugiwara, K. (2011); The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM). RFC 6116.

[4] Shen, C.; Schulzrinne, H. (2007); Measurement and Evaluation of ENUM Server Performance. *IEEE International Conf. on Communications, ICC 07*, 1967-1972.

[5] Shen, C.; Schulzrinne, H. (2006); Evaluation and Comparison of BIND, PDNS and Navitas as ENUM Server. Technical Report cucs-029-06, Columbia University.

[6] Tetenyi, I.; Szabo, G.; Kiss, A.; Toth, A. (2008); ENUM in everyday practice. Is it a dream or an opportunity?. *Infocommunication Journal*, LXIII:48-54.

[7] Rudinsky, J. (2009); Private ENUM based Number Portability Administrative System Evaluation. *International Conference on Ultra Modern Telecommunications Workshops, ICUMT 09*, 1-7.

[8] Lee, H.; Mun, Y. (2004); Performance Evaluation of ENUM Directory Service Design. *Lectures Notes in Computer Science*; 3038:1124-1130.

[9] Balakrichenan, S.; Bugnazet, T.; Becker, M. (2007); Studying ENUM Performance with Modeling and Simulation. *First Asia Int. Conf. on Modelling Simulation, AMS 07*; 82-87.

[10] da Mata, S.H.; Guardieiro, P.R.; Cardoso, A.; Lamounier, E.; Theodoro, L. (2011); Performance Study of ENUM Service Based in Modeling and Simulation. *8th Int. Conf. on Information, Communications and Signal Processing, ICICS 11*; 1-5, doi:10.1109/ICICS.2011.6173557.

[11] da Mata, S. H.; Magalhaes, J. M. H.; Cardoso, A.; Carvalho, H. A.; Guardieiro, P. R. (2013); Performance Comparison of ENUM Name Servers. *22nd International Conference on Computer Communication and Networks (ICCCN2013)*, 1-5.

[12] Nominum (2007). DNSPerf: DNS Performance Tool Manual.

[13] Lin, H. et al (2009); VoIP Signalling Performance Requirements and Expectations. Internet Draft.

[14] Vixie, P. et al (1997); Dynamic Updates in the Domain Name System (DNS UPDATE). RFC 2136.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):453-462, August, 2014.

Cloud Service Management System for Innovative Clusters. Application for North-West Region of Romania

D. Popescu, A. Dodescu, P. Filip

Daniela Popescu, Anca Dodescu*
University of Oradea
Economy Department
Romania, 410610 Oradea, University Street, 1
depopescu@uoradea.ro
*Corresponding author: adodescu@uoradea.ro

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

Abstract: In order to stimulate and optimize the organization and management of innovative clusters from value chain perspective and guide their planning activities towards a differentiation strategy in which cluster members cooperate, we propose a Cloud Service Management System (CSMS) that provides IT services for these innovative clusters companies that can be customized for both enterprises with the associated clusters.
Within such a system, actors begin to depend one on another and to take advantage of the local knowledge base. Each cluster is designed to have a different profile which will integrate all the companies mapped with it, with the objective of keeping the profile and data for each company. For the existing companies the idea is to migrate their services into the related cluster for integration within CSMS. Thus, our proposed CSMS will consider and meet different quality of services (QoS) parameters of each individual enterprise and service which will be included in specific Service Level Agreements (SLAs), after the negotiation between the cloud service provider and the CSMS. Realizing that technological progress is at the heart of regional development and decision-makers could support the development of technology clusters towards transforming them into regional innovative clusters, the application of our proposal aims to overcome existing bottlenecks in terms of business strategies and regional development policies in the North-West region of Romania.
Keywords: cloud computing, service oriented architecture, open cloud architecture, IT services, innovative clusters, supply chain management

1 Introduction

The paper proposes a *Cloud Service Management System (CSMS)* that provides IT services for the innovative clusters companies in order to develop collaborative mechanisms specific for innovative clusters aimed to solving the problems identified in the economic development of the North-West region of Romania.
Although in the past 20 years numerous studies have been conducted regarding the importance of innovative clusters for the regional economic development, the number of innovative clusters in Romania, in general, and in the North-West region, in particular, is surprisingly small, many of the existing clusters not being functional, due to lack of experience in organizing and managing

collaborative mechanisms specific for innovative clusters and insufficient and ineffective interconnection of companies in the same sector of activity through services and related technologies which enable an efficient group management and fostering innovation. Surprisingly small is also the number of networks of firms in Romania, the explanation being related to the lack of confidence of economic actors in the collaborative mechanisms and the lack of investments in building confidence at regional level, the lack of good governance practices of networks of firms that lead to bottom-up clustering as well as the lack of patterns of structured planning and development of innovative clusters.
In this context, the practical relevance of the problem sought to be solved through the present paper is the proposal of applying a Cloud Service Management System (CSMS) that provides IT services for innovative clusters which produces an effective group management, able to stimulate cooperation between cluster members in terms of the value chain on high value added niches, increasing competitiveness and innovation within the cluster and hence diversified smart specialization and regional development in order to increase the number and ensure long-term sustainability of innovative clusters in the North-West Region of Romania.
Proposing a functional model for the management of innovative clusters to compensate for functional limits of collaborative mechanisms within networks of companies and existing regional clusters, research involves an interdisciplinary cooperation of specialists in information technology, management and economics, the solutions targeted being positioned in the interference areas of the information technology with knowledge management and regional development.

2 Cloud Service Management System for Innovative Clusters

Given the advantages of cloud which allows the delivery of scalable resources on demand, we propose a Cloud Service Management System (CSMS) that provides IT services for the innovative clusters companies that can be customized for both enterprises with the associated clusters. In this sense, CSMS will be built within a private cloud with multiple clusters. We are focused to deliver broad range of cloud services as the ones described by David Linthicum in [2] due to their spread across industry. These are: Storage-as-a-service, Database-as-a-service, Information-as-a-service, Process-as-a-service, Application-as-a-service, Platform-as-a-service, Integration-as-a-service, Security-as-a-service, Management / Governance-as-a-service, Testing-as-a-service, Infrastructure-as-a-service. These services can be rent by customers depending on their needs and used as computing utilities. The advantage is the decreasement of the initial investment and that billing changes accordingly with the computing requirements for an individual or an organization changes, without incurring any additional cost.
Regarding the security of the Cloud Service Management System, we must point out that besides the stages of a Service Level agreement (SLA) content [1] [4], our services provided within each cluster of the CSMS must enhance availability, confidentiality, data integrity, control and audit. Within such a system, actors begin to depend one on another and to take advantage of the local knowledge base, if this interdependence causes a continuous flow of product and process innovations, diffusion of knowledge and collective learning processes at local or regional level, public-private partnerships being the synergistic factors that lead to innovative clusters.

Each cluster is designed to have a different profile (e.g. IT, automotive, apparel, tourism, health services, green energy etc.) which will integrate all the companies mapped with it, with the objective of keeping the profile of each company (Figure 1). For the existing companies the idea is to migrate their services into the related cluster for integration within CSMS. Eventually, CSMS will bring together all the involved services provided by the enterprise applications.
Moreover, in order to make the cloud computing approach to work as an efficient Cloud Ser-

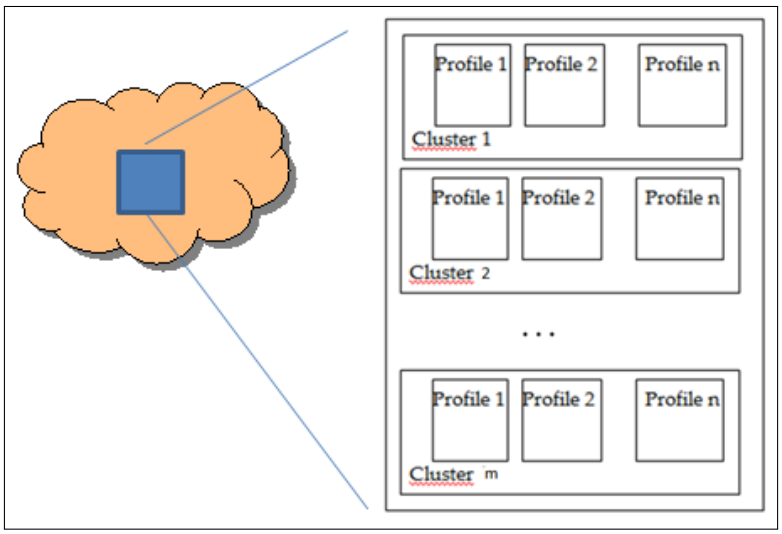


Figure 1: The CSMS/

vice Management for the innovative clusters, and to deliver the business agility and IT flexibility promised by Web Services, it is required the creation of a Service Oriented Environment (SOE) [3].Our approach is to deliver value to customers through the services provided using the SOE without implying any additional cost and risk from the customer perspective.
The services with CSMS are supposed to bring to customers: simplicity, flexibility maintainability, reusability and independence from technology. Services can be published, discovered and used in a technology neutral, standard form by using the web services protocols.
Thus, through a common interface, separate services for companies within each cluster are created and managed with the purpose of achieving the objectives of each company that has its own assets, employees, suppliers, partners (and information about them) or existing IT infrastructure [2] [6].
The huge increase in IT system interoperability that SOA can bring, not only at enterprise level, but also at the innovative cluster level is based on the use of smaller, modular services that have interface descriptions and contracts that ensure the business agility. These services are identified, defined and described in the context of the innovative cluster business activities and they are performed by the IT system and managed at the CSMS level. For each service is clearly set what it does and it is stipulated in a contract. The development of effective and flexible solutions is ensured by the use of SOA techniques like: service composition, discovery, message-based communication, and model-driven implementation [3].
SOA represents not only the architecture of services seen from a technology perspective, but also the policies, practices, and frameworks by which we ensure for the entire innovative cluster the right services are provided and consumed in order to ensure a best business outcome for the innovative cluster [2].
The software services used by the innovative cloud business operations are supported by a Cloud infrastructure that, together with the IT services, improves information flow within the companies and between the companies from the innovative cluster and all of them end the outside. The access to all of these services frequently involves a human-computer interface, often implemented as a web interface using portals (Figure 2), etc.

Thus, our proposed CSMS will consider and meet different quality of services (QoS) parameters of each individual enterprise and service which will be included in specific Service Level Agreements (SLAs), after the negotiation between the cloud service provider and the CSMS.

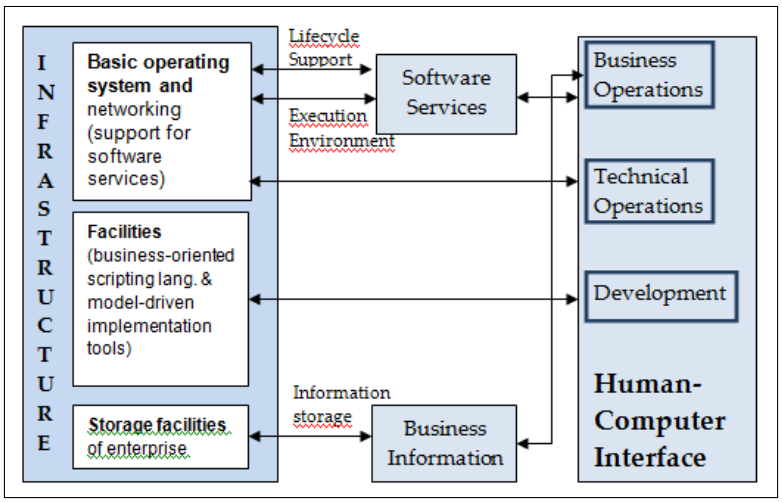


Figure 2: The Infrastructure (adapted from [3])/

Realizing that technological progress is at the heart of regional development and decision-makers could support the development of technology clusters towards transforming them into regional innovative clusters, the application of our proposal aims to overcome existing bottlenecks in terms of business strategies and regional development policies in the North-West region of Romania in order to stimulate and optimize the organization and management of innovative clusters from value chain perspective and guide their planning activities towards a differentiation strategy in which cluster members cooperate with high value added niches (smart diversified specialization) and, consequently, to create regional growth and development.
To get the ideal agile in collaborating environments (such as business) is needed that the IT infrastructure necessary to access the functionality by services be able to be configured by the user without the need to become experts in the field. A well suited solution can be the Cloud Computing Open Architecture (CCOA), proposed in [5].

Our CSMS solution for the management of services of the innovative cluster is developed based on the Cloud Computing Open Architecture, which we have to customize in order to ensure the services needed to migrate the companies services into the related cluster for integration within CSMS.
So, at the second level of the CCOA model [5] we develop the innovative clusters using an HPC infrastructure. Level 3, structured as SOA, will be used to define the profiles and the services for each company included in each cluster. Level 4 will be specific to each cluster, namely profile. Level 5 provides the granular services, while level 6 will have custom profile services for each cluster (Figure 3).

3 Innovative Clusters in North-West Region of Romania

There is currently a broad consensus to consider that economic agglomeration effects in a given location is the result of the balance of opposed agglomeration forces (centripetal and centrifugal), generally or in a particular field, branch of industry, assuming technological networking, mainly for efficiency and flexibility, or by propensity for innovation [7]. The identification of endogenous determinants of knowledge spillover and collective learning process within innovative clusters represent one of the most advanced theoretical perspectives on regional development [8]. Initially defined by Michael E. Porter as business strategy in a global economy context, as a source of competitive advantage for the participating companies [9], and later as "geographic

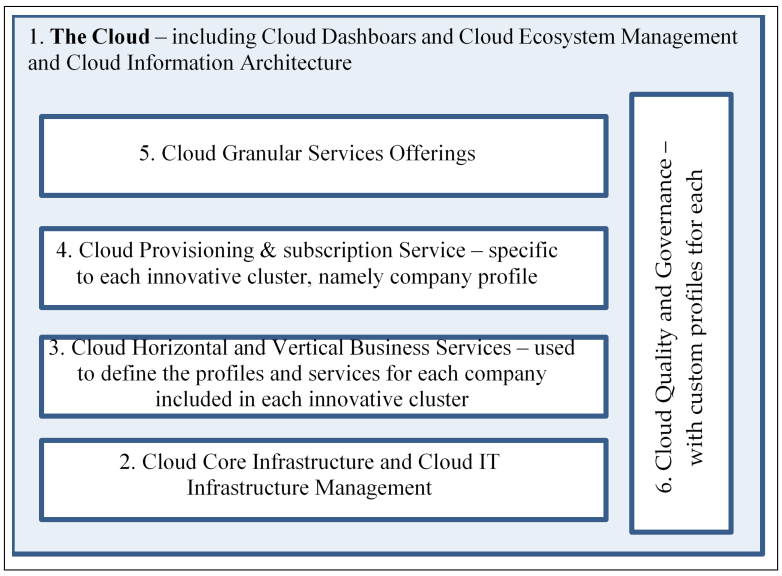


Figure 3: The CSMS Model/

concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries and associated institutions (eg universities, standardization agencies, professional associations) in a particular field that compete but also cooperate" [10], by encouraging the formation of some relational regional assets external to individual firms, but with major influence on the performance of their competitiveness [9], [12], the porterian cluster has turned from the most influential business strategy into the most popular regional development strategy. Innovation has become a critical factor for survival in a competitive spatial economy and also the factor that determines the direction and the rhythm of regional development. In this context, the propensity to form innovative clusters characterizes development policies at present, due to regional and local impact (providing a successful combination of infrastructure, highly skilled workforce and flexible corporations for development, offering well-paid and high quality jobs, have prospects for long-term growth and development, are dynamic and dominated by innovative SMEs because continuous research and development is vital for survival, it accelerates the technological and innovative process at local or regional level etc.) [13].

The linkages between firms and organizations that form an innovative clusters the institutional density at local level, the unsold interdependency network that go beyond market transactions, the mechanisms by which technological progress within the cluster turns into a collective learning system, have been the subject of numerous research based, especially, on case studies of successful technology clusters [14], the Cambridge technology cluster with the involvement of Cambridge University and science parks: Melbourne Science Park, Cambridge Science Park and St John's Innovation Park being frequently analyzed. Close working relationship between specialized SMEs, the involvement of higher education and research institutions, the free movement of highly skilled labor and availability of specialized consulting services within the cluster are considered the main explanatory mechanism for the success of these clusters.

In this context, the classical concept of network of companies, which designates a form of cooperation between companies legally independent, geographically dispersed, but with common interests in economic terms - has evolved into the concept of collaborative networks of innovative technological systems with respect to network configuration referring to coordinate cooperation between the various organizations that pursue a common goal: the development of a region; and the concept of technological cluster evolved into the "pôle de compétitivité" (in French ap-

proach) or innovative cluster facilitated by the availability of the nature of information technology to transform into a cumulative process and "institutional density" which facilitates cooperation and technological knowledge flow, leading to the accumulation of "strong local base of knowledge" and create a system like "innovative millieu" [15], [16], [17]. Within such a system, SMEs begin to depend one on another and to take advantage of the local knowledge base, if this interdependence does not stuck on the "old, competitive and unstable paths" [18], but causes a continuous flow of product and process innovations, diffusion of knowledge and collective learning processes at local or regional level, public-private partnerships being the synergistic factors that lead to innovative milieux [19], toward learning regions [20], creative regions, knowledge based regions etc. [8], [13].

Realizing that technological progress is at the heart of regional development and decision-makers could support the development of technology clusters towards transforming them into regional innovative clusters, learning environments and innovative regions or areas of knowledge, quickly penetrated regional growth and development policies of the EU Member States [15], [16], closely related with the pulses set in the Lisbon Strategy on "knowledge based economy" and later with the approach of the Territorial Agenda 2020 and the specific smart growth objectives of Europe 2020 strategy. The current financial and economic crisis is pressing even more to stimulate the innovation process and increase the role of regions in this context. The EU reports and strategic documents demonstrates that innovation policy should focus not only on research and development or science and technology, almost exclusively concentrated in a few regions of the EU, especially capital regions, but "to anticipate" appropriate policy mix for innovation covering all the dimensions of innovation, including those concerning organizational process, creativity, design, etc. and all the regions of Europe [21], [22], [23]], and the reality of a high concentration of technological capabilities in the developed center of the EU must change gradually and create space for regions and companies with a lower absorption capacity. But how to change this in favor of the less favored regions in conditions of economic crisis? The most common response is by stimulating both innovation and entrepreneurship through innovative regional clusters [22], [23]]. According to the 2013 European Cluster Observatory star rating [24], there are 92 clusters in Romania that have received at least one star depending to the extent to which clusters have achieved critical mass by employing measures in terms of size, specialization and focus [25].

In this context, the clustering potential of the North-West Region of Romania is sensitively high in the Figure 4 mentioned industries, where we represented the number of employees in the stars classified industries:

With a diverse economic structure, characterized by the shift from predominant development of strong labor intensive sectors to the services sector, particularly the intelligence intensive and high value-added industries, that require skilled and specialized labor force, North West Region of Romania has a certain degree of regional specialization for the following industries: furniture industry, electrical equipment, metallurgy, leather and footwear industry. for the following industries: furniture industry, electrical equipment, metallurgy, leather and footwear industry [26], according with Zaman's methodology for calculating apparent or revealed regional comparative advantage indicator, based on the top main regional trade exporters in the EU and between the EU extra trade [27].

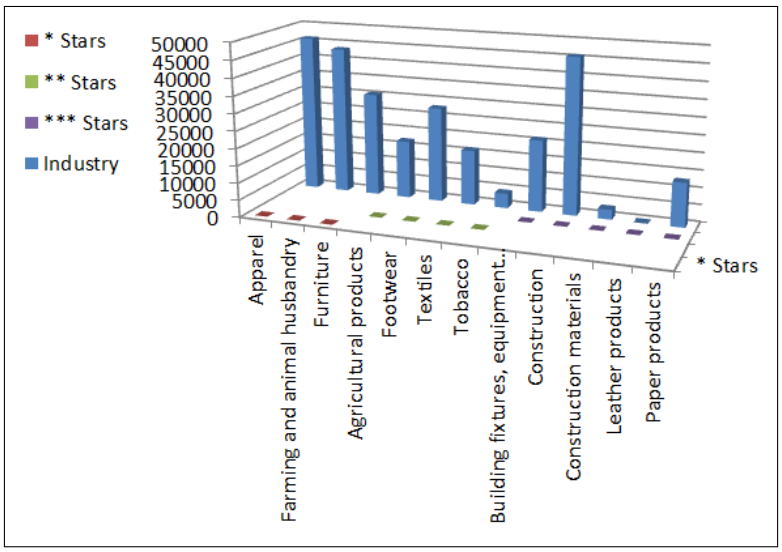


Figure 4: Star Clusters in North- West Vest Region of Romanial/

(Source: Star Clusters in Romania, Center for Strategy and Competitiveness, CSC Stockholm School of Economics, April 2011, http://www.clusterobservatory.eu/common/galleries/downloads/Star_clusters_Romania.pdf)

Regarding economic agglomeration, according with most recent regional development strategy 2014-2020 draft plan [26], in the North-West region there were 6 registered clusters, out of which only 4 are active as shown in Table below:

Table 1. Clusters situation in North-West Region of Romania

Cluster	Field of activity	Location	Information about association initiative under cluster structure and status
Romanian Water Cluster	water energy industry	Cluj-Napoca	Information not found - unfunctional
Geothermal Cluster	renewable energies, services in tourism	Oradea	Created in 2009 through CLUSTHERM (Creating a Central European Thermal Water Research Cluster) project, financed through FP7, Regions of Knowledge Programme (REGIONS-2007-2), Coordinator: INNOVA szak-alfld Regional Development and Innovation Nonprofit Ltd (HU), http://www.clustherm.eu/. - unfunctional
TREC - Transnational Renewable Energies Cluster	renewable energies	Cluj-Napoca	Created in 2012, through TREC (Transnational Renewable Energies Cluster) project, financed through SMART +, INTERREG IVC Programme, Coordinator: North-West Regional Development Agency, Romania, http://www.nord-vest.ro/Document_Files/Noutati/00001234/ciic6_1.%20TREC%20-%20concept,%20activitati,%20viziune.pdf. - active
Cluster Mobilier Transilvan	furniture	Cluj-Napoca	Created in 2012, bottom up, as a non-governmental organization through legal association of 18 firms from furniture industry, 3 catalyst organizations, 1 public institutions and 1 university, website: http://transylvanianfurniture.com/ - active
Cluj IT Cluster	IT & C	Cluj-Napoca	Created in 2012, bottom up, as a non-governmental organization through legal association of 28 IT & C companies, 6 public institutions and catalyst organizations, 4 universities, website: http://www.clujit.ro/ - active
Agro-Food-Ind Napoca	agro-food	Cluj-Napoca	Created in 2013, bottom up, as a non-governmental organization through legal association of 13 companies and associations from agro-food industry, 2 catalyst institutions, 3 public administration institutions, 2 universities; website: http://www.agrotransilvaniacluster.ro/ - active

Source: authors' processing

Looking at Figure 4 and Table 1, we should notice that, despite its sensitively high clustering

potential in apparel, farming and animal husbandry, agricultural products, footwear, textiles, tobacco, building fixtures, equipment and services, construction, construction materials, leather products, paper products [26], only furniture industry is covered with active cluster in the region. Surprising is the absence of a cluster in the apparel, footwear, textiles industries, despite existing industrial agglomerations and tradition in the region.
According to the Cluster Programme and Policy Stress Test for North-West Regional Development Agency Regional Innovation Strategy [28], the score of the North-West region is of 9 points out of 50 points, which places the region in the "yellow area" of the "league of clusters" traffic light, the minimum limit for compliance with the "green zone" being of 40 points. The weaknesses identified in the North West region of Romania after applying the policy stress test are related to the management approach of clusters and the internationalization of their activity, the malfunction integration of regional programs for clusters in the general context of industrial policy and economic development policy, the critical points being related to the implementation, monitoring, evaluation and impact of regional programs for clusters and the absence of a national policy for clusters.

4 Conclusions

The analysis of economic agglomerations in North-West Region of Romania shows the predominance of economic agglomerations based on simple agglomeration economies (especially, industrial parks) and high difficulties in the implementation and ensuring sustainability on long term of economic agglomerations based on economies of dynamic agglomeration (especially, innovative clusters) and highlights the need of instruments to improve the management approach of clusters and to stimulate the innovative process, and also collective learning processes through cooperation and relationality. In this context, the proposed Cloud Service Management System model aims to represent an important support for the development of existing clusters and to find solutions for the functionality and sustainability of new clusters in the North-West region for valuing high regional clustering potential. In this respect, broad range of cloud services that can be rent by firms from same industry, aligning cluster members on value chain. A number of services can be adapted to the cluster specificity: project planning, marketing planning, strategy, business plan, KPI's system etc.; knowledge management, open virtual library, e-learning, working groups, application procurement, legislation Info, Match-making, funding opportunities Portal, common EU projects, etc. Within such a system, actors begin to depend one on another and to take advantage of the local knowledge base, if this interdependence causes a continuous flow of product and process innovations, diffusion of knowledge and collective learning processes at local or regional level, public-private partnerships being the synergistic factors that lead to innovative clusters. The proposed model will help to tackle the functionality and long-term sustainability issues of innovative clusters in the North-West region of Romania and will represent the base of an integrated network that will bring together human resources trained within the project in interdisciplinary applied research and SMEs networks from the North-West region of Romania and will contribute decisively to the accumulation of the "critical mass" of resources for research, development and innovation for increasing regional competitiveness.

Through the development of some model for structured planning and development of innovative clusters, our future steps will be focused on testing of model on a pilot cluster in order to stimulate learning, relationships and cooperation within innovative clusters, the development of the following innovative services: methodologies and methods for planning and structured development of innovative clusters, handbook for the management of innovative clusters, system of integrated software for managing clusters and knowledge, and fostering cooperation within an innovative cluster.

Bibliography

[1] Lonea, A.M., Tianfield, H. and Popescu, D.E. (2013). Identity Management for Cloud Computing, *New Concepts and Applications in Soft Computing*, 417: 175-199.

[2] Linthicum, D. S. (2009); *Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide*, Addison-Wesley Information Technology Series.

[3] Open Group, *SOA and Enterprise Architecture*, Available online: http://opengroup.org/ [14.12.2013].

[4] Lonea, A.M., Popescu, D.E. and Prostean, O. (2012); The Overall Process Taken by Enterprises to Manage the IaaS Cloud Services, *Proceedings of the 6th European Conference on Information Management and Evaluation*, University College, Cork, Ireland, 13-14 September 2012.

[5] Vitek, A. J. & Morris, M.N, (2012); Service Oriented Cloud Computing Architectures, *UMM CSci Senior Seminar Conference*, Available online: https://wiki.umn.edu/ [07.11.2013].

[6] Entreprise Concept, Available online: http://enterprise-concept.com/ [05.12.2013].

[7] Sövell, Ö., Lindqvist, G. & Ketels, Ch. (2003); *The Cluster Initiative Greenbook*, Stockholm: Bromma tryk AB.

[8] Capello, R. & Nijkamp, P. (eds.) (2009); *Handbook of Regional Growth and Development Theories*, Edward Elgar Publishing.

[9] Porter, M. E. (1990); *The Competitive Advantage of Nations*, New York: Free Press.

[10] Porter, M.E. (1998); *On Competition*, Boston: Harvard Business School Press.

[11] Porter, M.E. (1998); Clusters and the new economics of competition, *Harvard Business Review*, November-December, 77-90.

[12] McCann, P. (2001); *Urban and Regional Economics*, Oxford: Oxford University Press.

[13] Armstrong, H. & Taylor, J. (2000); *Regional Economics and Policy*, Third Edition. Oxford: Blackwell.

[14] Karlsson, C. (2008); *Handbook of Research on Clusters*, Edward Elgar Publishing House.

[15] Ketels, Ch. & Svell, A. (2006); *Innovation Clusters in the 10 New Member States of the European Union*, European Communities, Available online: http://www.isc.hbs.edu/ [22.07.2013].

[16] European Commission (2007); *Innovation Clusters in Europe. A Statistical Analysis and Overview of Current Policy Support*, Europe Innova / PRo Inno Europe paper no. 5, DG Enterprise and Industry report.

[17] Hamdouch, A. (2008); Conceptualizing Innovation Clusters and Networks, *International Conference: Innovation Networks*, Tacoma-Seattle, Washington, USA, May 14-16, 2008, Available online: http://rrien.univ-littoral.fr/ [22.07.2013].

[18] Armstrong, H. & Taylor, J. (2000); *Regional Economics and Policy*, Third Edition. Oxford: Blackwell.

[19] Camagni, R. (ed.) (1991); *Innovation Networks: Spatial Perspectives*, London: Belhaven Press.

[20] Morgan, K. (1997); The Learning Region: Institutions, Innovation and Regional Renewal, *Regional Studies*, 31(5);491-503.

[21] European Commission (2011); *The urban and regional dimension of Europe 2020.Seventh progress report on economic, social and territorial cohesion*, Report from the Commission, Available online: http://ec.europa.eu/ [24.09.2013].

[22] European Commission (2012); *State of the Innovation Union Report 2012 - Accelerating change*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - SWD(2013) 75 final, Available online: http://eur-lex.europa.eu/ [24.09.2013].

[23] European Commission (2013); *Innovation Union progress at country level 2013*, Available online: http://ec.europa.eu/ [24.09.2013].

[24] European Cluster Observatory (2013);Available online: http://www.clusterobservatory.eu/ [18.02.2014].

[25] European Cluster Observatory (2011); *Star Clusters in Romania, Center for Strategy and Competitiveness*, CSC Stockholm School of Economics, April 2011, Available online: http://www.clusterobservatory.eu/ [13.11.2013].

[26] Romanian North-West Regional Development Agency (2013); 2014-2020 Regional Development Plan (draft, September 2013), Available online: http://www.nord-vest.ro/ [18.02.2014].

[27] Zaman, Gh. (2010); *Sistemul de indicatori ai competitivitatii romanesti in contextul dezvoltarii durabile, communication for AFER Summer School*, University of Oradea, September 2010.

[28] Lammer-Gamp, K. (2013); *Results Cluster Programme and Policy Stress for North-West Regional Development Agency Regional Innovation Strategy*, European Secretariat for Cluster Analysis, Berlin

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):463-470, August, 2014.

A Conceptual Architecture of Ontology Based KM System for Failure Mode and Effects Analysis

Z. Rehman, S. Kifor

Zobia Rehman*
1. Faculty of Engineering and Management
Lucian Blaga University of Sibiu, Romania
2. Department of Computer Science
Comsats Institute of Information Technology, Pakistan
*Corresponding author: zobia.rehman@gmail.com

Stefania Kifor
Lucian Blaga University of Sibiu, Romania
kifors@yahoo.com

Abstract: Failure Mode and Effects Analysis (FMEA) is a systematic method for procedure analyses and risk assessment. It is a structured way to identify potential failure modes of a product or process, probability of their occurrence, and their overall effects. The basic purpose of this analysis is to mitigate the risk and the impact associated to a failure by planning and prioritizing actions to make a product or a process robust to failure. Effective manufacturing and improved quality products are the fruits of successful implementation of FMEA. During this activity valuable knowledge is generated which turns into product or process quality and efficiency. If this knowledge can be shared and reused then it would be helpful in early identification of failure points and their troubleshooting, and will also help the quality management to get decision support in time. But integration and reuse of this knowledge is difficult because there are number of challenges e.g., unavailability of unified criteria of FMEA knowledge, lack of semantic organization, natural language text based description of knowledge, most of the times FMEA is started from scratch instead of using existing knowledge that makes it incomplete for larger systems, and above all its success depends on the knowledge which is stored in the brains of perfectionists in the form of experience which may or may not be available anytime anywhere. In this article we are proposing an Information and Communication Technology (ICT) based solution to preserve, reuse, and share the valuable knowledge produced during FMEA. In proposed system existing knowledge available in repositories and experts head will be gathered and stored in a knowledge base using an ontology, and at the time of need this knowledge base will be inferred to make decisions in order to mitigate the probable risks. Ontology based approaches are best suited for the knowledge management systems, in which human experts are required to model and analyze their expertise in order to feed them in a conceptual knowledge base for its preservation and reuse.
Keywords: Knowledge Management (KM), ontology, Failure Mode and Effects Analysis (FMEA).

1 Introduction

Knowledge is often defined as "justified true belief". Facts, information, and skills acquired by practical experience or education are also called knowledge. Knowledge is the information combined with experience, context, interpretation, and reflection. It is a high-value form of information that is ready to apply to decisions and actions [3]. In this era knowledge is the core asset of any organization. It is important for the organizations to know what they know. Nowadays sustainability of an organization depends on what it knows, how efficiently it utilizes what it knows, and how quickly it acquires and utilizes new knowledge [3], and in other words success of

any organization relies on the effective management of its knowledge. Organizational knowledge can reside in the minds of experts and experienced employs of organizations, files in cabinets, or in huge volumes of databases on servers. Since ages human beings and organizations have been creating, acquiring, communicating, and re-utilizing their knowledge. Knowledge management can be defined as doing what is needed to get the most out of the knowledge resources [6]. It is also the collection of activities of knowledge gathering, structuring, refinement, and dissemination. Nowadays ICT is helping to make these activities more effective for organizations. Knowledge management systems have been and are being designed and developed to facilitate organizations in managing their knowledge and making it useful in the absence of concerned experts. But knowledge management systems are very difficult to build. It is difficult to find experts who are willing to share their knowledge, and allow experiencing how they solve problems. It is difficult to gather, compile and refine the knowledge across the organization from person to person, documenting the way how decisions are reached, and structuring it to represent on machine.

A project risk is defined as an uncertain and undesired event or condition which affects the objectives of a project in positive or negative way [15].Risk management is an endeavor to identify and manage internal events and external threats that can affect the likelihood of a project's success. It identifies risk events, plans how to minimize the consequences or impact of the risk event, lists all anticipated tasks that can be completed in order to avoid undesired events, and builds contingency plans to handle such events after their occurrence. It is a proactive approach to reduce surprises and negative consequences. It provides better control over the future, makes a project manager ready to take the advantage of appropriate risks; and increases the probability of meeting project objectives within budget and time [8]. Successful project management is based on effective risk planning and management. Risk management helps a project manager to actively handle all expected and unexpected risks on projects. There are various other factors that affect a project's success but inadequate risk management increases the possibility of project failure.

Risk management process is based on four major actions; risk identification, risk assessment, risk response development, and risk response control [1]. In risk identification phase all probable risks are considered which can potentially make a process or product fail to achieve its objectives; brain storming, problem identification, and risk profiling are effective approaches for this phase. In risk assessment phase, the value of risk associated to a product or process is determined by calculating the magnitude and probability of an undesired event. There are different tools for risk assessment, e.g. scenario analysis for event probability and impact, risk assessment matrix, FMEA, probability analysis; and Semi-quantitative scenario analysis. Risk response development phase is responsible for deciding an appropriate response for assessed risk. This response could be mitigation, avoidance, transfer, or retention of the risk. To control risk, response strategies are designed and executed, events are monitored, contingency plans are initiated and new risks are identified; all these tasks are completed in risk response control phase.

FMEA is an effective systematic technique for failure analysis during product life cycle. Its purpose is to identify all potential failure modes related to a system, product or process in advance, so that experts can think about their remedies and can take actions to reduce the probability of their occurrence or severity of their impacts [18]. For the very first time FMEA was used in the 1960's in the aerospace industry during the Apollo missions. In 1974, the US Navy developed MIL-STD-1629 about the use of FMEA. In the late 1970's, the automotive industry started using FMEA and they realized about the advantages of using this tool to reduce risks in order to improve quality of the product [18]. Now this inductive reasoning analysis is in practice in almost every field of engineering and technology and especially it has become a preliminary step of quality engineering, reliability engineering, and safety engineering. FMEA report outlines the

findings which have been developed from worksheets. The findings focus on the failure modes which can have significant effects on the system and categorize them as catastrophic, critical, down to minimal, nuisance value, or etc. An FMEA covering the complete system includes its constituent FMEAs by a reviewing and analyzing the interfaces between the subsystems. During the analysis some failure modes are difficult to assess, so FMEA is responsible for the assessment of such failure modes by executing practical tests and generating their results. In time prepared comprehensive FMEA provides technical support in identification of weak points in a system, reduces the cost of modification along the product life cycle [2][4].Organizations spend a lot of effort and high cost to apply FMEA method but still the knowledge they acquire during this analysis is neither reused nor shared. Major reason is that the failure modes are not semantically organized. Their interpretation varies from person to person and from situation to situation. Even there is a possibility that a team prepared an FMEA in one situation, might not be able to apply the same to some other situation. Also reusing of this information becomes imprecise and unproductive because of its larger size and its production again and again [12]. So an IT based solution is needed to store, share, reuse, and infer the knowledge produced by FMEA. Artificial Intelligence (AI) has been supportive in not only storing and manipulating data but also in acquiring, manipulating, and representing knowledge. Knowledge based systems based on the concept of ontologies and expert systems have been popular for addressing knowledge management problems. In this article we will focus on the ontology based knowledge management systems which have been formulated for FMEA. We also propose a conceptual model of an ontology based knowledge management system for FMEA, which we believe is able to solve all above mentioned problems. Ontology is an axiomatic theory, designed for a specific domain composed of objects, and is made explicit by means of a formal specific language [21].

2 Related Work

There are different technologies which have been helpful for applying knowledge, but AI played major role in enabling these systems to work. AI systems are based on the belief that intelligence and knowledge are closely related to each other. Knowledge is associated with the human's capability to manipulate cognitive symbols, whereas human intelligence deals with our ability to learn and communicate our understandings in order to solve problems. Human beings are born with a certain degree of intelligence, by using this they learn and acquire new knowledge. There are some AI systems which are known as knowledge based systems and their purpose is to mimic problem solving capabilities of the experts of a specific domain. These intelligent systems and their associated technologies assist us in every phase of knowledge management [6]. Four AI techniques have been widely used to build knowledge management systems, which are Case Based Reasoning (CBR), rule based expert systems, fuzzy logic, and ontologies. Ontologies are found better than rest of the three, as some disadvantages are associated with rule-based systems (fuzzy or crisp). For example in many situations, a large number of rules are required to represent the domain knowledge, and it could make the inference mechanism quite complex and time consuming, as it adds difficulty in coding, verifying, validating, and maintaining these rules. Moreover they do not have any ability to learn [6]. Whereas the major disadvantage associated with CBR is that all cases are needed to be stored in case library with additional care about their references and attributes, efficient methods are needed to search the cases and identify their search attributes, and they do not provided well presented information to their users [11].
Purpose of ontology development is to provide semantically shared domain knowledge for intelligent inferencing, since ontologies are famous for domains like semantic web, system engineering, bio-medical informatics,software engineering, e-commerce, knowledge engineering, knowl-

edge management, artificial intelligence, natural language processing, etc. Ontologies play a distinct role in information retrieval and information systems, they better serve as system models and help in storing and retrieving knowledge semantically. Ontology engineering is a sub-field of knowledge engineering. Its aim is to make knowledge explicit, which is hidden in software applications, business procedures, and in enterprise repositories [16]. While constructing an ontology for any domain, knowledge acquisition is the major bottleneck, so in the knowledge management systems. But there are some automated knowledge acquisition techniques which can be applied e.g., linguistic techniques to acquire knowledge from unstructured text documents, machine learning techniques to acquire knowledge from structured documents, crawling through web, and requesting experts for required knowledge according to a template [14].

Some work has been found in which ontologies have been helpful in managing knowledge produced by FMEA. Authors in [12] applied ontology based knowledge retrieval approach for distributed knowledge sharing and reuse in the PFMEA domain. Using this strategy the authors were able to construct a complex search patterns, combining many complex roles of PFMEA, and connecting them to a particular instance. This system helped them to establish actions with adequate knowledge retrieval and semantic organization. In [5] the authors realized that traditional way (text based) to represent knowledge produced and used during FMEA, makes it difficult to reuse. That's why they developed an ontology, based on the standard ISO-15926 to improve the representation of knowledge used and produced during FMEA. In [13], authors are trying for lead free soldered, for its accomplishment they need to carry an FMEA to find all critical failures associated to this process and their impact on the quality of lead free soldered joints. They suppose that ontology is suitable to store the knowledge produced during FMEA as it offers the common understanding of the domain of their focus and its procedures; and it is also a computationally processed model to represent knowledge. Authors in [4] believe that FMEA is highly effective method to prevent the failure of a system in early phases of its development but a lot of effort is needed to develop an FMEA and it is difficult to earn the advantages immediately. To get more advantages from FMEA it is needed to shorten its development process by reusing the knowledge gathered in already developed FMEA. Knowledge produced by FMEA is in natural language, that's why systematized components, functions and failure modes are not explicit and are hard to reuse. Most of the times the meanings of such natural language documents, depend on the understanding of the team who prepare these and may not be clearly interpret-able for a different one. Authors found that ontology could help to integrate the related knowledge by providing uniform semantic knowledge of a domain for solving a fore mentioned problems. In [7] authors tried to transform scattered knowledge of extended functional model to FMEA sheets using ontology. Functional model included the knowledge of functional structure and faulty behavior. They developed the ontology by identifying the knowledge models in various forms and mapped the ontology by identifying correspondence relations between the concepts of two ontologies. Authors were also trying to transform this knowledge to FTA. Authors in [20] found that FMEA knowledge structure lacks unified standards because of this it cannot be collected during manufacturing process. Because of this for many organizations FMEA is not an effective and feasible method for improvement in manufacturing process. They proposed a system to extract and convert FMEA existing knowledge to develop a systematic, structured ontology-based FMEA knowledge management system. This system was helpful for intelligent inquiry and reasoning from FMEA knowledge in order to make decisions timely, to handle quality issues in manufacturing process. Similarly [4] [9][19]used ontology to smooth the progress of FMEA proceedings.

They all have done a lot for reusing the knowledge gathered during FMEA using ontologies for different industries but still some very important things are missing from the scene. For the very first time [10] presented an approach to bind FMEA and ontologies together but this model

lacks the facility of inference because according to authors ontology is just a conceptual model without having any rules. Authors in [4] focused on the deficiencies left by [10] by combining knowledge management and quality management systems to reuse the existing FMEA knowledge easily but according to them this work still misses a commonsense ontology to provide the parts of standardized and functional taxonomy. Authors in [13] proposed a method for better FMEA procedure representation using ontologies for lead free soldering but still his work is under research. Moreover no specific ontology is found to address the FMEA knowledge sharing and reuse for automotive domain.
In this article we are proposing a conceptual architecture for ontology based knowledge management system specifically for automotive domain.

3 Proposed Conceptual Architecture

Working of knowledge management systems is based on four major actions; Knowledge gathering, knowledge organization and structuring, knowledge refinement, and knowledge distribution. Figure 1, depicts a conceptual architecture for a knowledge management system and in the following subsections it is described in terms of four major knowledge management actions.

3.1 Knowledge Gathering

Knowledge gathering is the set of activities which are responsible for acquisition of the knowledge to be managed. According to the given model it will be accomplished by gathering existing FMEA reports completed by experts and also by gathering the rationales and heuristics of the knowledge of domain experts.

3.2 Knowledge Organization and Structuring

From information science perspective an ontology is described as formal description of a domain, in terms of instances and relationship between those instances. According to the given model it will describe the basic structure for the instances which will be stored in the knowledge base. Knowledge to be structured and stored will be provided to ontology populator after its extraction from existing FMEA reports and document files composed on the heuristics and rationales of experts. Ontology populator will learn the structure of instances and their relationships from ontology and will populate the knowledge in the knowledge base. To extract knowledge from existing reports some Natural Language Processing techniques will be applied, but depending on the accuracy of the techniques some human intervention would be needed.

3.3 Knowledge Refinement

Once the existence knowledge is gathered the next phase is about its refinement. For refinement of this knowledge some experts will be invited to query the system for specific problems of domain. Purpose of this querying would be to validate the knowledge in the knowledge base that if it is correct or enough to solve all probable problems expected in the domain. These experts will be authorized to update the knowledge base according to their level of expertise. They can add new instances and also can update the existing ones.

3.4 Knowledge Distribution

Once the knowledge base is updated with appropriate and refined knowledge, it is ready to assist its users via browser interface. Inference engine will depict the rationales of experts and will provide intelligent conclusions after inferring through knowledge base. This conceptual architecture is equally useful in multi-user environment, where number of authenticated users can interact with the system via internet.

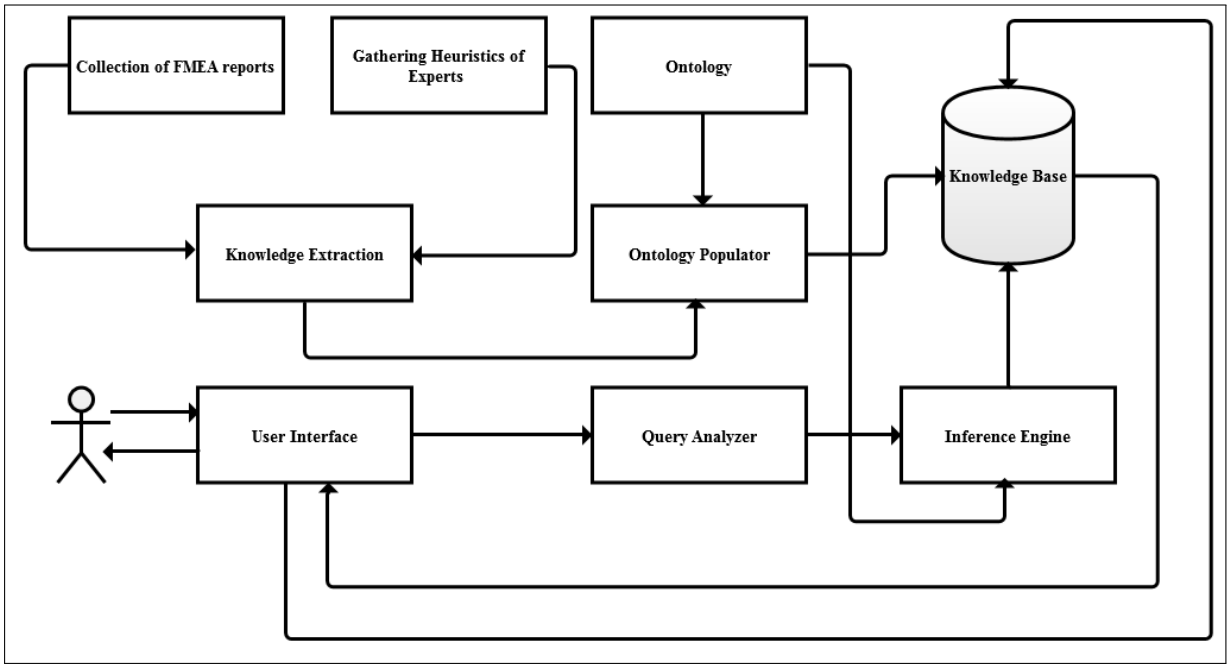


Figure 1: Conceptual Architecture of Ontology Based Knowledge Management System for Failure Mode and Effects Analysis (FMEA)

4 Required Tools and Techniques

Collection of FMEA reports would be in a formal way by requesting the organizations to provide these. Heuristics of experts over FMEA can be gathered by interviewing them about a given scenario and cross validating their views. These views can be stored in machine readable documents. Once the basic knowledge is gathered, GATE (General Architecture for Text Engineering) can be used to extract required knowledge, It is an open source infrastructure to develop and implement software components to process human languages, it helps in keeping data separate from application. We will use Protege to create ontology and its instances. Protege is an open source software tool to develop domain models and knowledge based systems with ontologies. It provides both of the main ways to model ontologies e.g., frame based and OWL. It provides full reasoning support to compute the inferred ontology class hierarchy and for consistency checking. OWL-Lite can be used to get full support of existing Description Logic reasoners for inference engine. Jena can be used to query the ontology and to update its instances.

5 Conclusion

Strategic management of any organization revolves around the risk management. Good risk management focuses on the identification and treatment of the risk in time, to increase the probability of the success and decrease the probability of the failure in achieving overall objectives of an organization. FMEA is a systematic way to identify probable failure modes and their overall effects on a system, if this analysis is timely completed, it helps to reduce the probability of risk occurrence. In this article we proposed a conceptual architecture for ontology based knowledge management system to eliminate the problems regarding poor management of the knowledge produced during FMEA, to timely generate FMEA reports even in the absence of domain experts, and to reuse and share existing FMEA knowledge. Ontologies are like repositories to organize formal knowledge and information represented by common vocabulary of a language to support reasoning, they are becoming the core component for many applications e.g., knowledge portals, knowledge management and integration systems, and e-commerce applications. Ontologies are helpful in exchange of knowledge, reusing existing knowledge, for reasoning and inferring on existing knowledge [17]. As ontologies use common understanding of concepts of a domain this is why FMEA knowledge stored in ontology will be general, comprehensive, machine readable and explicit. In future we will materialize this conceptual model for automotive domain.

Bibliography

[1] Carbone, T. A.; Tippett, D.; (2004); Project Risk Management Using the Project Risk FMEA, *Engineering Management Journal*, 16, 28-35.

[2] Chang, K.H.; Ching, C.H.; (2010); A risk assessment methodology using intuitionistic, *International Journal of Systems Science*, 1457–1471.

[3] Davenport , T. ; Prusak, L.; (1998); *Working knowledge*, Boston, MA : Harvard Business School Press.

[4] Dittmann, L. et al (2004); Performing FMEA using ontologies, *18th International workshop on qualitative reasoning*, 209–216.

[5] Ebrahimipour, V. et al (2010); An ontology approach to support FMEA studies, *Elsevier Expert Systems with Applications*, 671-677.

[6] Fernandez, B.I.; Saberwal, R.; (2010); *Knowledge Management Systems and Processes*, M.E. Sharpe.

[7] Koji, Y. et al (2005); Ontology-based transformation from an extended functional model to FMEA, *International conference on engineering design*, Melbourne.

[8] Larson, E. W.; Gray, C. F.; (2011); *Project Management: The Managerial Process* (5th Edition ed.), McGraw Hill.

[9] Laaroussi, A. et al (2007); Ontology-aided FMEA for construction products, *J. IEEE Press*, 189-194.

[10] LEE, C.F.; (2001); Using FMEA models and ontologies to build diagnostic models, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 281-293.

[11] Mansouri, D.; Hamdi-Cherif, A.; (2011); Ontology-oriented case-based reasoning (CBR) approach for trainings adaptive delivery, *15th WSEAS Int. Conf. on Computers (CSCC2011)*, 328-333.

[12] Mikos, W.L. et al (2011); A system for distributed sharing and reuse of design and manufacturing, *Elsevier Journal of Manufacturing Systems*, 133-143.

[13] Molhanec, M. et al (2010); The Ontology based FMEA of Lead Free Soldering Process, *International Spring Seminar on Electronics Technology, ISSE.*

[14] Natalya, F. N.; Deborah, L. M.; (2001); Ontology Development 101: A Guide to Creating Your First Ontology, `http://protege.stanford.edu/publications/ontology_development/ontology101.html`.

[15] PMI; (2000); Project Management Body of Knowledge Guide PMBOK. Pennsaylvania: project management body of knowledge guide PMBOK.

[16] Pouchard, Line; Ivezic, Nenad; Schlenoff, Craig (March 2000); Ontology Engineering for Distributed Collaboration in Manufacturing, *Proceedings of the AIS2000 conference.*

[17] Sharman, R. et al (2007); *Ontologies A hand book of principles, concepts, and applications in information systems*, Springer.

[18] Stamatis, D.H.; (2003); *Failure mode and effect analysis: FMEA from theory to execution*, USA: ASQ Quality Press.

[19] Wang, Y.M. et al (2009); Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean, *Elsevier Expert Systems with Applications*, 1195-1207.

[20] Xiuxu, Z.; Yuming, Z.; (2012); Application Research of Ontology-enabled Process FMEA Knowledge Management Method, *Int. J. Intelligent Systems and Applications*, 34-40.

[21] Zuniga, G.L.; (2001); Ontology: Its transformation from philosophy to information system, *Proceedings of international conference on formal ontology*, 187-197.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):471-481, August, 2014.

Networked Predictive Fuzzy Control of Systems with Forward Channel Delays Based on a Linear Model Predictor

S.W. Tong, D.W. Qian, G.P. Liu

Shiwen Tong*
College of Automation
Beijing Union University
Beijing 100101, China
* Corresponding author: shiwen.tong@buu.edu.cn

Dianwei Qian
School of Control and Computer Engineering
North China Electric Power University
Beijing 102206, China
dianwei.qian@ncepu.edu.cn

Guo-Ping Liu
1.School of Engineering
University of South Wales
Portypridd CF37 1DL, UK
2.CTGT Center
Harbin Institute of Technology
Harbin 150001, China
guoping.liu@southwales.ac.uk

Abstract: This paper presents a novel networked control framework, using fuzzy logic control, for systems with network delays which are known to greatly weaken the control performance of the controlled system. To deal with the network delays, the predicted differences between the desired future set-points and the predicted outputs from a model predictor are utilized as the inputs of a fuzzy controller, thus a series of future control actions are generated. By selecting the appropriated control sequence in the plant side, the network delays are compensated. The simulative results demonstrate that the proposed method can obviously reduce the effect of network delays, and improve the system dynamic performance.
Keywords: networked control, model prediction, fuzzy control, delay compensation.

1 Introduction

The emergence of the network technology has changed the communication architecture of control systems from traditional point-to-point to current common bus. Sensors, actuators and controllers are connected through network, formed a feedback control system (namely, networked control system). This architecture has injected fresh blood to the classic and modern control theories and also arises higher challenges to the controller design at the same time. On the one hand, the introduction of the network to the control system brings many advantages such as low cost, easy maintenance and high reliability. On the other hand, the unavoidable time delay, data dropout and other complicated phenomenon existing in the network should be considered. In recent years, networked control theory and technology have become an important and hot research area. Scholars from different countries have made a lot of breakthroughs in the networked control [1]- [8].

Network delay has become one of the most concerned issues in the networked control system. Because the network delays can dramatically degrade the control performance of the systems

even makes the systems unstable. From recent published literatures, it can be seen that the treatment of the network delays can be summarized in the following: The first one is to integrate delay information into the controller design by designing a robust controller to decrease the effect of time delay [3], [6]- [8]; The second one is to estimate delay information in the backward or forward channel by using reason-rule-table, average-value or delay windows (DW) method [9,10]; The third one is to eliminate the network delays in the return path by using a cascade control structure with P (proportion) control in the inner loop and fuzzy adaptive control in the outer loop [4]; As we all known, one of the obvious characteristics of the networked control system is that the communication networks can transmit a packet of data simultaneously. This feature provides another solution to compensate for network delays in the forward channel [11,12]. Based on it, Liu *et al*. [1, 2] proposed networked predictive control (NPC), using the strings of future control actions, to compensate for the forward channel delays. In this paper, we try to design a networked controller by using fuzzy control theories. Different from networked predictive control method, we separate the model prediction from the controller design. The function of the model predictor is just to produce future predicted outputs. A fuzzy controller is designed to generate a series of future control sequence based on the errors between the desired future outputs and the model predicted outputs. Then the strings of future control actions are packed and sent to the plant side through the communication channel. Thus, the effect of delays in the forward channel is lessened by using a delay compensator in the plant side.

Predictive control and fuzzy control are powerful tools. They have been used in the design of the networked controller [1]- [5], [7, 8]. Some researchers have connected the prediction to the fuzzy control and have proposed 'predictive fuzzy control' method [13, 14]. They use 'future error' and 'future error change' as inputs of the fuzzy controller to produce the control actions at the current time. To my knowledge, the combination of networked control system and the predictive fuzzy control has not been reported except one paper written by us in 2007 [15]. We call it 'networked predictive fuzzy control' with abbreviation NPFC. The core idea of the NPFC is producing 'future control actions' by fuzzy controller design according to the 'future error' and 'future error change' from a model predictor. Then the network delays can be compensated by choosing the 'future control actions'.

This paper is organized as follows: The architecture of networked predictive fuzzy control (NPFC) is firstly presented. Secondly, a model predictor based on Diophantine equation is proposed. Thirdly, the fuzzy controller using 'future errors' and 'future error changes' as inputs to derive 'future control actions' is designed. Then the delay compensation mechanism is discussed and the method is implemented in a servo control system. Finally, the conclusions are drawn in Section 4.

2 Design of networked control systems

2.1 Structure of Networked Predictive Fuzzy Control Systems

The networked predictive fuzzy control system as shown in Fig. 1 mainly consists of three key parts: the model predictor, the fuzzy controller and the delay compensator. The model predictor is used to predict future outputs of the controlled system $y(t|t), y(t+1|t), \cdots, y(t+N-1|t)$ according to the delayed output $y(t-1)$ of the controlled system in the backward channel and the control actions $u(t-d-1), u(t-d-2), \cdots, u(t-d-n_b)$ in the past. The errors $e(t|t), e(t+1|t), \cdots, e(t+N-1|t)$ between the desired future outputs $r(t|t), r(t+1|t), \cdots, r(t+N-1|t)$ and the predictive future outputs $y(t|t), y(t+1|t), \cdots, y(t+N-1|t)$ of the controlled system can be used to design a fuzzy controller to produce the future control sequences $u(t|t), u(t+$

$1|t), \cdots, u(t+N_u-1|t)$. Then the future control sequences are packed and sent to the plant side through network. In the plant side, a delay compensator is used to compensate for the forward network delays by selecting appropriate control sequence.

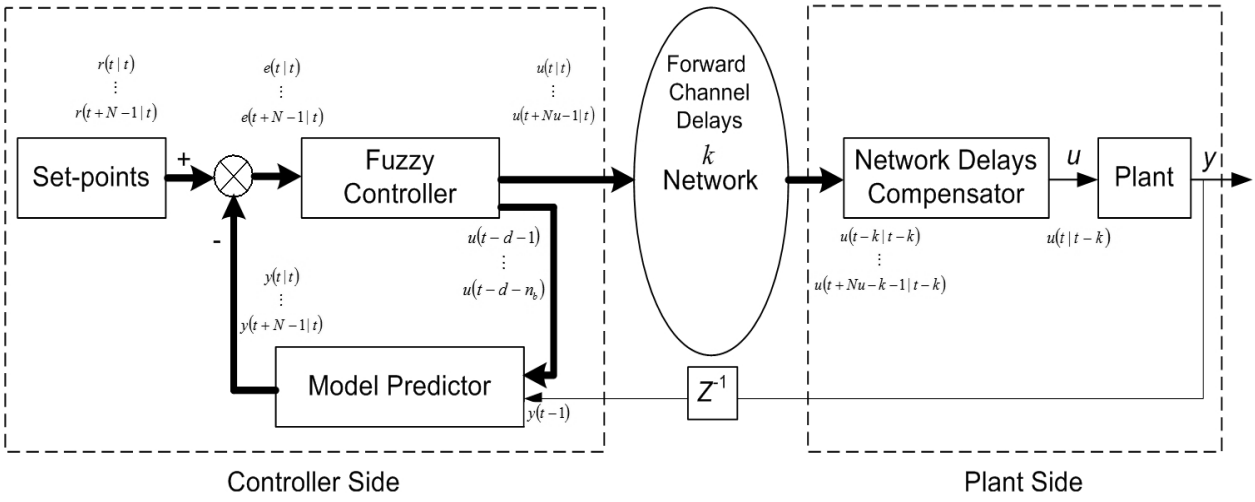


Figure 1: Structure of Networked Predictive Fuzzy Control System

2.2 Model Predictor

The function of the model predictor is to produce a series of future outputs of the controlled object according to the delayed output in the backward channel and the control actions in the past. Models, no matter linear or nonlinear model, of such functions can all be the predictor. To simplified the issue and focus on the research work step by step, we only consider forward channel delays in this paper. This consideration is reasonable in some cases. For example, in a network that the transmission velocity in the backward channel is faster than the forward channel thus the delays in the backward channel can be neglected. Therefore a linear model predictor based on the Diophantine equation is proposed.

Consider a single-input and single-output process with the following form

$$A(z^{-1})y(t) = z^{-d}B(z^{-1})u(t) \tag{1}$$

where,

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}$$

Introducing a Diophantine equation to derive the model predictor.

$$\Delta A(z^{-1})E_i(z^{-1}) + z^{-i}F_i(z^{-1}) = 1 \tag{2}$$

where, $E_i(z^{-1})$ is of order $i-1$ and $F_i(z^{-1})$ is of order n_a.

$$E_i(z^{-1}) = 1 + \sum_{j=1}^{i-1} e_{i,j} z^{-j}, F_i(z^{-1}) = \sum_{j=0}^{n_a} f_{i,j} z^{-j}.$$

Define N as predictive horizon, N_m as model control horizon, from (1) and (2), the predicted value $Y_p(t+1)$ of the controlled system can be obtained.

$$Y_p(t+1) = P\Delta U(t-d) + Q\Delta U(t-d-1) + Fy(t-1) \tag{3}$$

where,

$$P = \begin{bmatrix} g_0 & 0 & \cdots & 0 \\ g_1 & g_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{N_m-1} & \cdots & \cdots & g_0 \\ \vdots & \vdots & \cdots & \vdots \\ g_{N-1} & g_{N-2} & \cdots & g_{N-N_m} \end{bmatrix}$$

$$Y_p(t+1) = [y_p(t|t), y_p(t+1|t), \cdots, y_p(t+N-1|t)]^T$$

$$\Delta U(t-d) = [\Delta u(t-d), \Delta u(t-d+1), \cdots, \Delta u(t-d+N_m-1)]^T$$

$$\Delta U(t-d-1) = [\Delta u(t-d-n_b), \Delta u(t-d-n_b+1), \cdots, \Delta u(t-d-1)]^T$$

$$F = [F_1(z^{-1}), F_2(z^{-1}), \cdots, F_N(z^{-1})]^T$$

$$Q = \begin{bmatrix} g_{n_b} & \cdots & g_1 \\ \vdots & \cdots & \vdots \\ g_{n_b+N-1} & \cdots & g_N \end{bmatrix}$$

Let $G_i(z^{-1}) = B(z^{-1})E_i(z^{-1}) = g_0 + g_1 z^{-1} + \cdots + g_{n_b+i-1} z^{-(n_b+i-1)}$, we can construct matrix P and Q by solving the Diophantine equation.

It should be noted that the item $\Delta U(t-d) = [\Delta u(t-d), \Delta u(t-d+1), \cdots, \Delta u(t-d+N_m-1)]^T$ in equation (3) can not be obtained at the current time t. We assume $\Delta u(t-d+i) = 0, i = 0, 1, \cdots, N_m - 1$. This assumption is reasonable. On the one hand, the dynamic response of the future time can be reflected by using the past control actions and the previous process outputs. On the other hand, the item $y(t-1)$ is corresponding to the correction of the predicted model. That is $y(t-1) = y_m(t-1) + e(t-1)$, where $y_m(t-1)$ is the model output and $e(t-1)$ is the error between the process output and the model output.

2.3 Fuzzy Controller

The successful implementation of networked predictive fuzzy control relies on an assurance that the fuzzy controller must calculate all the pre-defined control actions in every sample time. So the fuzzy control algorithm should be simple and suitable to the real-time control. A real-time simplified fuzzy controller [16] is used to generate the future control actions. Unlike the ordinary fuzzy control, which uses error and error change at the present time to produce the current control action, the networked predictive fuzzy controller uses 'future error' and 'future error change' to derive 'future control actions'. The ordinary fuzzy control can be regarded as a special condition of the networked fuzzy control when the predictive step is equal to zero. This networked predictive fuzzy controller has two-input one-output. One input is the error e between the desired future output and the predicted output. The other one is the change of the error ec. The output of the fuzzy controller is the change of the future control action Δu. The membership functions of e and ec are adopted as triangular forms and the membership function of the output Δu is adopted discrete form as Fig. 2 shown.

The 'Simplification' of the fuzzy controller relays on the inference process. For the typical two-input-one-output fuzzy controller, only four control rules are excited at each cycle time with

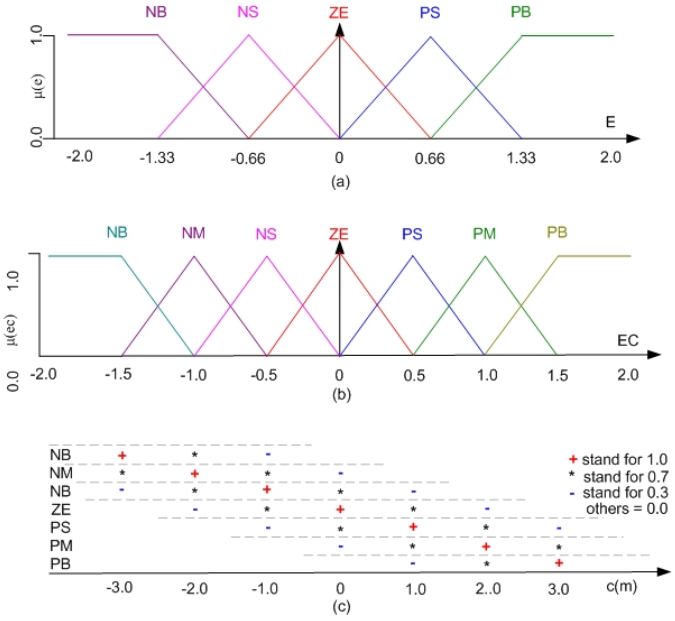


Figure 2: Membership functions of e, ec and Δu. (a) is the membership function of input e, (b) is the membership function of input ec, (c) is the discrete form membership function of output Δu.

the form 'If e is L_e and ec is L_{ec}, then Δu is $L_{\Delta u}$', where L is the linguistic variables PB, PM, PS, ZE, NS, NM, NB. Due to the characteristic of the triangular membership function, e is at most belong to two membership functions μ_e^i and μ_e^{i+1}, ec is at most belong to two membership functions μ_{ec}^j and μ_{ec}^{j+1}, thus Δu has 2×2 combinations, that is four control rules:

if e is $L_e^{(i)}$ and if ec is $L_{ec}^{(j)}$ then Δu is $L_{\Delta u}^{(i,j)}$
if e is $L_e^{(i)}$ and if ec is $L_{ec}^{(j+1)}$ then Δu is $L_{\Delta u}^{(i,j+1)}$
if e is $L_e^{(i+1)}$ and if ec is $L_{ec}^{(j)}$ then Δu is $L_{\Delta u}^{(i+1,j)}$
if e is $L_e^{(i+1)}$ and if ec is $L_{ec}^{(j+1)}$ then Δu is $L_{\Delta u}^{(i+1,j+1)}$

From Fig. 2(c) shown, output Δu is adopted discrete form membership function. It is assumed that output domain has been divided into $c_m(k), m = 1, 2, \cdots, n$. Do minimum and maximum operator, calculate

$$\begin{aligned}
\widetilde{\mu}_{L_{\Delta u}^{(i,j)}}^{(m)} &= \wedge(\mu_{L_e^{(i)}}, \mu_{L_{ec}^{(j)}}, \mu_{L_{\Delta u}^{(i,j)}}^{(m)}) \\
\widetilde{\mu}_{L_{\Delta u}^{(i,j+1)}}^{(m)} &= \wedge(\mu_{L_e^{(i)}}, \mu_{L_{ec}^{(j+1)}}, \mu_{L_{\Delta u}^{(i,j+1)}}^{(m)}) \\
\widetilde{\mu}_{L_{\Delta u}^{(i+1,j)}}^{(m)} &= \wedge(\mu_{L_e^{(i+1)}}, \mu_{L_{ec}^{(j)}}, \mu_{L_{\Delta u}^{(i+1,j)}}^{(m)}) \\
\widetilde{\mu}_{L_{\Delta u}^{(i+1,j+1)}}^{(m)} &= \wedge(\mu_{L_e^{(i+1)}}, \mu_{L_{ec}^{(j+1)}}, \mu_{L_{\Delta u}^{(i+1,j+1)}}^{(m)})
\end{aligned} \tag{4}$$

and

$$\widetilde{\widetilde{\mu}}_{\Delta u}^{(m)} = \vee(\widetilde{\mu}_{L_{\Delta u}^{(i,j)}}^{(m)}, \widetilde{\mu}_{L_{\Delta u}^{(i,j+1)}}^{(m)}, \widetilde{\mu}_{L_{\Delta u}^{(i+1,j)}}^{(m)}, \widetilde{\mu}_{L_{\Delta u}^{(i+1,j+1)}}^{(m)}) \tag{5}$$

where $m = 1, 2, \cdots, n$

It is not easy to directly get the inference rules of the future; however, the typical dynamic of the second order system can be obtained ahead of time. Fig. 3 presents the phase plane of

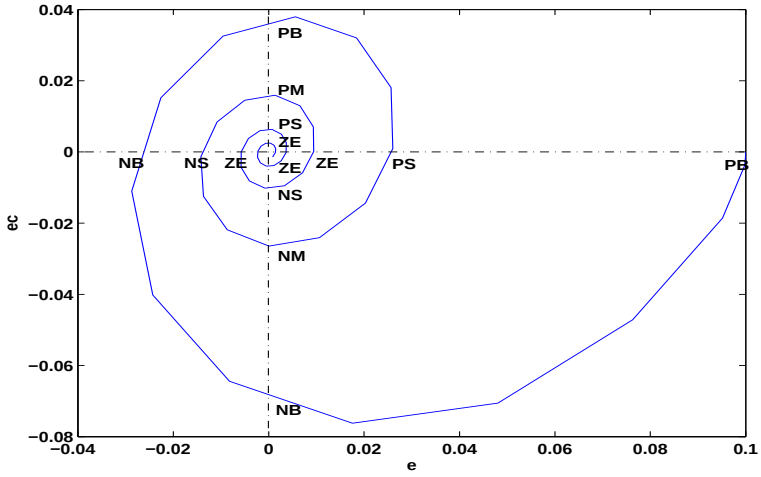


Figure 3: Phase plane and fuzzy control rules analysis of typical second order linear system.

the typical second order linear system with the x-axis standing for variable e and y-axis standing for variable ec. From this figure, it can be seen that the points in x-axis and y-axis are crucial. If these points are controlled very well, the control performance is guaranteed. So the main inference rules are deduced as shown in Table 1.

Table 1: Main control rules of the fuzzy controller in NPFC

No.	Control rules
1	if $e = PB$ and $ec = ZE$ then $\Delta u = PB$
2	if $e = ZE$ and $ec = NB$ then $\Delta u = NB$
3	if $e = NB$ and $ec = ZE$ then $\Delta u = NB$
4	if $e = ZE$ and $ec = PB$ then $\Delta u = PB$
5	if $e = PS$ and $ec = ZE$ then $\Delta u = PS$
6	if $e = ZE$ and $ec = NM$ then $\Delta u = NM$
7	if $e = NS$ and $ec = ZE$ then $\Delta u = NS$
8	if $e = ZE$ and $ec = PM$ then $\Delta u = PM$
9	if $e = ZE$ and $ec = NS$ then $\Delta u = NS$
10	if $e = ZE$ and $ec = PS$ then $\Delta u = PS$
11	if $e = ZE$ and $ec = ZE$ then $\Delta u = ZE$

Thus, the incremental control actions

$$\Delta u(t+\eta) = \frac{\sum_{m=1}^{n} \tilde{\mu}_{\Delta u}^{(m)} * c_m}{\sum_{m=1}^{n} \tilde{\mu}_{\Delta u}^{(m)}} \tag{6}$$

and the predicted control actions

$$u(t+\eta) = u(t+\eta-1) + \Delta u(t+\eta), \eta \text{ is integer and } 0 \le \eta \le N_u - 1 \tag{7}$$

can be given, where N_u is the control horizon. If e is the error and ec is the error change at present time, then Δu is the incremental control action at present time. If e and ec are the values of the future time, then the future incremental control actions can be derived.

As paper [17]- [19] presents, variable domain implemented in the fuzzy control can greatly improve the control accuracy, which has successfully been applied to the control of a quadruple inverted pendulum [20] and the wing rock phenomenon [21]. This is achieved by the domain contraction and expansion. Domain contraction is equivalent to the increase of the control rules. Although the real-time algorithm has considered only 11 rules, together with the variable domain strategy, the fuzzy controller can acquire satisfactory control performance. The varies of domain can be achieved by multiplying flex factors $\alpha(e)$, $\beta(ec)$ and $\gamma(e, ec)$ of domains E, EC and ΔU, respectively. There are many different forms of flex factors. In this paper, we adopt the exponential form of functions:

$$\begin{aligned} \alpha(e) &= \left[\frac{|e|}{E}\right]^{\tau_1}, 0 < \tau_1 < 1 \\ \beta(ec) &= \left[\frac{|ec|}{EC}\right]^{\tau_2}, 0 < \tau_2 < 1 \\ \gamma(e, ec) &= \left[\left[\frac{|e|}{E}\right]^{\tau_1} \left[\frac{|ec|}{EC}\right]^{\tau_2}\right]^{\tau_3}, 0 < \tau_1, \tau_2, \tau_3 < 1 \end{aligned} \tag{8}$$

To summarized, the networked predictive fuzzy controller has eleven parameters to be designed. Four parameters are related to the model predictor. They are the order of the predictive model n_a and n_b, the predictive horizon N, and the model control horizon N_m. Seven parameters are belong to the networked fuzzy controller. They are the control horizon N_u, the scaling gains K_e, K_{ec}, $K_\Delta u$ of error e, error change ec and incremental control action Δu, and the variable domain parameters τ_1, τ_2 and τ_3. The program steps for the networked predictive fuzzy control plus variable domain strategy are summarized below:

Step 1: Use equation (3), calculate the future outputs $y_p(t|t), y_p(t+1|t), \cdots, y_p(t+N-1|t)$ of the controlled system according to the delayed output of the feedback channel and the previous control actions.

Step 2: Calculate the differences between the desired future outputs $r(t|t), r(t+1|t), \cdots, r(t+N-1|t)$ and the model predicted values $y_p(t|t), y_p(t+1|t), \cdots, y_p(t+N-1|t)$ to get $e(t|t), e(t+1|t), \cdots, e(t+N-1|t)$ and $ec(t|t), ec(t+1|t), \cdots, ec(t+N-1|t)$.

Step 3: Adjust input and output domain using equation (8) in terms of $e(t+\eta)$ and $ec(t+\eta)$.

Step 4: Calculate membership functions of input $e(t+\eta)$ and $ec(t+\eta)$ and output $\Delta u(t+\eta)$.

Step 5: Use minimum-maximum inference method [see equation (4) and (5)].

Step 6: Calculate the predicted control actions $u(t+\eta)$ using equation (6) and (7).

Step 7: Let $\eta = 0$ to $N_u - 1$, repeat step 3-6.

Step 8: Send the control actions $\Delta u(t)$,$\Delta u(t+1)$,$\cdots$,$\Delta u(t+N_u-1)$ with a packet to the plant side.

Step 9: Select the control action $u(t|t-k)$ and add to the controlled process.

Step 10: In the next sample time, repeat step 1-9.

2.4 Network Delays Compensation

It is assumed that the network communication delay in the forward channel is not greater than the length of the predicted control horizon. To make use of the 'packet transmission' characteristic of the network, a string of future control actions which contain $u(t), u(t+1), \cdots u(t+N_u-1)$ at sample time t are sent to the plant side at the same time. Then the control value from the latest control sequence available on the plant side is chosen as a control input of the plant to compensate for the forward channel delay. For example, if the latest control sequence

on the plant side is

$$\begin{bmatrix} u(t-k|t-k) \\ u(t-k+1|t-k) \\ \vdots \\ u(t-k+Nu-1|t-k) \end{bmatrix}$$

Then the output selected control signal will be

$$u(t) = u(t|t-k)$$

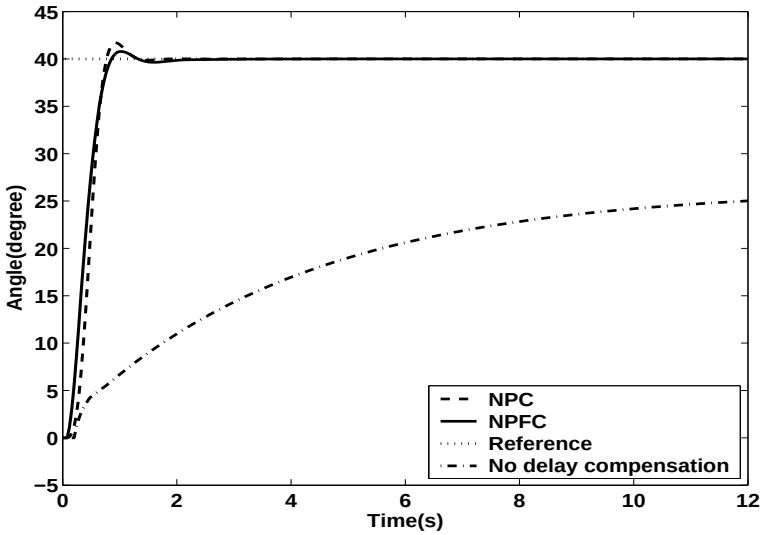


Figure 4: The step responses of NPC and NPFC with 1-step forward channel delay.

3 Simulations

Consider the servo control system as reference [2] shown. The system model with discrete form is as follows:

$$G(z^{-1}) = \frac{-0.00886z^{-1} + 1.268227z^{-2}}{1 - 1.66168z^{-1} + 0.6631z^{-2}} \tag{9}$$

where sample time is 0.04 second.

Suppose there is one step delay in the forward channel, following Section 2.2 and Section 2.3, we design a linear model predictor and a real-time fuzzy controller. The parameters of the model predictor are: $n_a = 2$, $n_b = 2$, $N = 12$, $N_m = 10$. The parameters of the networked fuzzy controller are: $N_u = 10$, $K_e = 0.00125$, $K_{ec} = 0.02$, $K_{\Delta u} = 0.6$, $\tau_1 = 0.1$, $\tau_2 = 0.1$ and $\tau_3 = 0.01$. The NPC parameters are set to $N = 12$, $N_u = 10$, $\rho = 1500$. Fig. 4 shows the control performance of NPFC and NPC. The dot line is the set-point. The solid line stands for the NPFC method. The dash line stands for NPC method. The dash-dot line stands for the NFPC method without delay compensation. From the figure, it can be seen that the NPFC can be regulated better than the NPC in control performance with rapid dynamic and small overshot. The delay compensation mechanism is very effective.

Suppose the case that six step delays exist in the forward channel. The NPFC controller parameters are adjusted as: $K_e = 0.0004$, $K_{ec} = 0.008$, $K_{\Delta u} = 0.08$, and the NPC parameters

are set to $N = 25$, $N_u = 10$, $\rho = 100000$. To testify the control performance of the networked predictive fuzzy control method, the results of the NPC and the NPFC are presented in Fig. 5. Through model prediction, fuzzy controller design and delay compensation, the NPFC presents very obviously better performance than NPC method. The rising time of NPFC is about 1.1 seconds while 1.5 seconds for NPC method. Moreover, NPC has 3.75% overshot while NPFC has nearly no overshot. When NPFC method not considers delay compensation, static errors can be seen in Fig. 5. On the contrary, the dynamic response reaches steady state after 1.4 seconds when the delay compensator is acting.

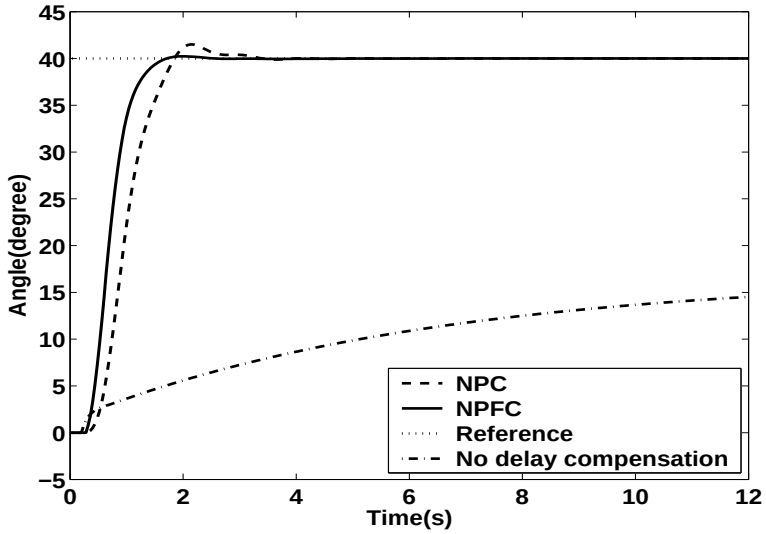


Figure 5: The step responses of NPC and NPFC with 6-step forward channel delay.

4 Conclusions

This paper proposes a network predictive fuzzy controller structure. By solving the Diophantine equation, the linear model predictor generates a series of predictive outputs of the controlled system according to the outputs and the control sequence in the past time. The errors between the desired future outputs and the predictive outputs from the linear model predictor and the error change are used to design a real-time fuzzy controller. So a series of future control sequence is produced in the controller side. By selecting the appropriate control sequence in the plant side, the delays in the forward channel are compensated. Because NPFC has more parameters which can be regulated, the control performance can be adjusted better than NPC method.

Acknowledgments

This work was supported in part by the Beijing Municipal Excellent Talents Funded Project (Grant No. 2013D005022000003), the Korea Foundation for Advanced Studies Fellowship to Dianwei Qian, and the National Natural Science Foundation of China (Grant No. 61273104 and 61333003).

Bibliography

[1] Liu G. P.; Rees D.; Chai S. C.; Nie X. Y. (2005); Design, simulation and implementation of netowrked predictive control systems, *Measurement and Control*, ISSN 2029-2940, 38:17-21.

[2] Liu G. P.; Mu J. X.; Rees D.; Chai S. C. (2006); Design and stability analysis of networked control systems with random communication time delay using the modified MPC, *International Journal of Control*, ISSN 0020-7179, 79(4):288-297.

[3] Mahmoud M. S.; Saif A. A. (2012); Robust quantized approach to fuzzy networked control systems, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, ISSN 2156-3357, 2(1):71-81.

[4] Du F.; Qian Q. Q. (2008); The research of heterogeneous networked control systems based on modify smith predictor and fuzzy adaptive control, *IEEE International Conference on Fuzzy Systems*.

[5] Tang P. L.; De S. C. W. (2006); Compensation for transmission delays in an ethernet-based control network using variable-horizon predictive control, *IEEE Transactions on Control Systems Technology*, ISSN 1063-6536, 14(4):707-718.

[6] Jia X. C.; Zhang D. W.; Zheng L. H.; Zheng N. N. (2008); Modeling and stabilization for a class of nonlinear networked control systems: A T-S fuzzy approach, *Progress in Natural Science*, ISSN 1002-0071, 18(8):1031-1037.

[7] Jiang X. F.; Han Q. L. (2008); On designing fuzzy controllers for a class of nonlinear networked control systems, *IEEE Transactions on Fuzzy System*, ISSN 1063-6706, 16(4):1050-1060.

[8] Hajebi P.; Almodarresi S. M. T. (2012); Online adaptive fuzzy logic controller using neural network for Networked Control Systems, *International Conference on Advanced Communication Technology*.

[9] Ren C. Q.; Wu P. D.; Wang X. F.; Ma S. Y.; Chen Z. L. (2002); A study on the forecast arithmatic of hydraulic telecontrol system based on internet, *Journal of Beijng Institute of Technology*, ISSN 1004-0579, 22(1):85-89. (In Chinese)

[10] Zhen W.; Xie J. Y. (2002); On-line delay-evaluation control for networked control systems, *IEEE Conference on Decision and Control*.

[11] Zhang Y. Y.; Zhang J. L.; Luo X. Y.; Guan X. P. (2013); Faults detection for networked control systems via predictive control, *International Journal of Automation and Computing*, ISSN 1476-8186, 10(3):173-180.

[12] Tang X. M.; Ding B. C. (2012); Design of networked control systems with bounded arbitrary time delays, *International Journal of Automation and Computing*, ISSN 1476-8186, 9(2):182-190.

[13] Li P. F.; Yan X. P.; Qiu L. P.; Zhou Q. Y.(2009); Study on predictive fuzzy control of great inertia system based on grey model, *2009 Second International Conference on Intelligent Computation Technology and Automation*.

[14] Hu J. Q.; Rose E.(1997); Predictive fuzzy control applied to the sinter strand process, *Control Engineering Practice*, ISSN 0967-0661, 5(2):247-252.

[15] Tong S. W.; Liu G. P.(2007); Design and Simulation of Fuel Cell Networked Predictive Fuzzy Control Systems, *Proceedings of the 26th Chinese Control Conference.*

[16] Tong S. W.; Liu G. P.(2008); Real-time simplified variable domain fuzzy control of pem fuel cell flow systems, *European Journal of Control*, ISSN 0947-3580, 14(3):223-233.

[17] Li H. X. (1995); To see the success of fuzzy logic from mathematical essence of fuzzy control, *Fuzzy Systems and Mathematics*, ISSN 1001-7402, 9(4):1-14. (In Chinese)

[18] Oh S. Y.; Park D. J. (1995); Self-tuning controller with variable universe of discourse, *IEEE International Conference on Systems, Man and Cybernetics.*

[19] Li H. X. (1999); Variable universe adpative fuzzy controller, *Science in China (Series E)*, ISSN 1006-9321, 20(1):32-42.

[20] Li H. X.; Miao Z. H.; Wang J. Y. (2002); Variable universe adaptive fuzzy control on the quadruple inverted pendulum, *Science in China (Series E)*, ISSN 1006-9321, 45(2):213-224.

[21] Liu Z. L.; Su C. Y.; Svoboda J. (2004); Control of wing rock phenomenon with a variable universe fuzzy controller, *Proceeding of the American Control Conference.*

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):482-496, August, 2014.

Observer-Based Non-Fragile Passive Control for Uncertain Nonlinear Sampled-Data System with Time-Delay

S.G. Wang, Y.S. Li, Y. Wu

Shigang Wang
School of Mechanical & Electric Engineering
Heilongjiang University
No.74 XueFu Road, Harbin, 150080, P. R. China
wsg6363@gmail.com

Yingsong Li*
College of Information and Communications Engineering
Harbin Engineering University
No.145 NanTong Street, Harbin, 150001, P. R. China
*Corresponding author: liyingsong82@gmail.com

Yi Wu
School of Information Science and Technology
Heilongjiang University
No.74 XueFu Road, Harbin, 150080, P. R. China
wy51cn@yahoo.com.cn

Abstract: The problem of observer-based passive control for uncertain nonlinear sampled-data systems with time delay is investigated by using non-fragile passive control. Attention is focused on the design of a non-fragile passive observer and a controller which guarantees the passivity of the closed-loop system for all admissible uncertainties. A sufficient condition for passivity and asymptotic stability of the combined system is derived via linear matrix inequality (LMI). Finally, a simulation example is presented to show the validity and advantages of the proposed method.
Keywords: Uncertain sampled-data system, Time-delay systems, State observer, Non-fragile passive control, Linear matrix inequality

1 Introduction

In the past few years, sampled-data systems are widely encountered in the area of control theory and control engineering, such as welding process, aerospace, signal processing, earthquake prediction, due to its character as continuous control plant and discrete controller [1]. Since time delay and inherent nonlinearity often occurs and causes serious deterioration of the stability of various engineering system, considerable research has been done and numerous results have been obtained relating to the analysis and synthesis of uncertain nonlinear sampled-data systems with time-delay, see e.g. [2–7]. Among these results, non-fragile passive control problem have attracted particular attention. Passivity is part of a broader and more general theory of dissipativity and non-fragility is a scheme of solving robustness of controller and observer themselves [8–11], they maintain the system's internal stability. Because non-fragile passive control has attractive features such as fast response, good transient response and insensitivity to variations in system parameters and external disturbance [12–16], which is likely to be an effective technique of control for uncertain nonlinear sampled-data system with time-delay.

On the other hand, all above works are based on an implicit assumption that the states are all known. However, this unrealistic assumption is not always verified. and hence, the construction of the unmeasured states through the knowledge of the system's inputs and outputs still an unavoidable task to solve any desired control problem [17–21]. However, to our knowledge, there

have been few results in literature of any investigation for observer-based non-fragile passivity uncertain nonlinear sampled-data system with time-delay.

The above situation is exactly what concerns and interests us. a novel approach of non-fragile control combined with passive control is proposed for stabilizing a class of uncertain nonlinear systems with time-delay. By utilizing a non-fragile state observer, a novel control law is established such that the resulting closed-loop system is strictly passive. A sufficient condition for the passivity and asymptotic stability of the augmented system is derived via LMI. Finally, an example is simulated to illustrated the advantage of the proposed method.

2 Problem Statement and Preliminaries

Consider the plant of uncertain nonlinear sampled-data system with time-delay described by

$$\begin{cases} \dot{x}(t) = (A_0 + \Delta A_0)x(t) + (A_1 + \Delta A_1)x(t-\tau) + B_0u(t) + f(x,u,t) + B_1\omega(t) \\ y(t) = C_1x(t) + H_2u(t) \\ z(t) = C_2x(t) + H_3\omega(t) \\ x(t) = x_0\ ,\ t \in [-\tau,\ 0] \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state, and $u(t) \in \mathbb{R}^m$ is the control input, $y(t) \in \mathbb{R}^p$ is regulated output, $z(t) \in \mathbb{R}^q$ is measured output, $\omega(t) \in \mathbb{R}^r$ is the external disturbance input that belongs to $\mathbb{L}_2[0,\infty]$, $A_0, A_1, B_0, B_1, C_1, C_2, H_2, H_3$ are known real constant matrices of appropriate dimension, $\Delta A_0, \Delta A_1$, are uncertain matrices. f is the uncertain nonlinear function vector, $f(0,0,t_0) = 0$, and f satisfies the Lipschitz condition.

Assumption 2.1. *The continuous plant is time-driven with a constant sampling period* $h(h > 0)$.

Discretizing system (1) in one period, we can obtain the discrete state equation of the plant of sampled-data system

$$\begin{cases} x(k+1) = (G_0 + \Delta G_0)x(k) + (G_1 + \Delta G_1)x(k-d) + H_0u(k) + \bar{f}(x_k, x_{k-d}, k) + H_1\omega(k) \\ y(k) = C_1x(k) + H_2u(k) \\ z(k) = C_2x(k) + H_3\omega(k) \\ x(k) = x_0\ ,\ k \le 0 \end{cases} \tag{2}$$

where

$G_0 = e^{A_0h}\ ,\ G_1 = \int_0^h e^{A_0(h-w)}dwA_1$

$H_0 = \int_0^h e^{A_0(h-w)}dwB_0\ ,\ H_1 = \int_0^h e^{A_0(h-w)}dwB_1$

$\bar{f}(x,u,k) = \int_0^h e^{A_0w}dwf(x,u,t)$

$\bar{f}(\hat{x},u,k) = \int_0^h e^{A_0w}dwf(\hat{x},u,t)$

Consider non-fragile observer described by

$$\begin{cases} \hat{x}(k+1) = G_0\hat{x}(k) + G_1\hat{x}(k-d) + H_0u(k) + (L + \Delta L)(y(k) - \hat{y}(k)) + \bar{f}(\hat{x}_k, x_{k-d}, k) \\ \hat{y}(k) = C_1\hat{x}(k) + H_2u(k) \end{cases} \tag{3}$$

where $\hat{x}(k) \in \mathbb{R}^n$ is state of observer, and L is observer gain, ΔG_0, ΔG_1, are uncertain matrices and ΔL are observer gain perturbation which are assumed to be of the following form:

$$[\Delta G_0 \quad \Delta G_1 \quad \Delta L] = M_0 F(k)[E_0 \quad E_1 \quad E_2] \tag{4}$$

On the other hand non-fragile controller described by

$$u(k) = (K + \Delta K)\hat{x}(k) \tag{5}$$

where K is controller gain, ΔK represents corresponding gain perturbation, and generally, there exist the following two classes of perturbation in ΔK:

Type1: ΔK is of the additive form:

$$\Delta K = M_0 F(k) E_3 \tag{6}$$

Type2: ΔK is of the multiplicative form:

$$\Delta K = M_0 F(k) E_4 K \tag{7}$$

where M_0, E_0, E_1, E_2, E_3, and E_4 are real matrices with appropriate dimension and $F(k) \in \mathbb{R}^{k\times l}$ is an unknown time-varying matrix function satisfying

$$F^{\mathrm{T}}(k)F(k) \le I$$

Assumption 2.2. *$\bar{f}(x_k, x_{k-d}, k), \bar{f}(\hat{x}_k, \hat{x}_{k-d}, k)$ satisfies the quadratic inequality in the domains of continuity, that is*

$$\bar{f}^{\mathrm{T}}(x_k, x_{k-d}, k)\bar{f}(x_k, x_{k-d}, k) \le \delta_1^2 x^{\mathrm{T}}(k) M_1^{\mathrm{T}} M_1 x(k) + \delta_{d1}^2 x^{\mathrm{T}}(k-d) M_{d1}^{\mathrm{T}} M_{d1} x(k-d) \tag{8}$$

Let $\xi_1^{\mathrm{T}} = \left[\begin{array}{cccc} e^{\mathrm{T}}(k) & \hat{x}^{\mathrm{T}}(k) & e^{\mathrm{T}}(k-d) & \hat{x}^{\mathrm{T}}(k-d) & \bar{f}^{\mathrm{T}}(x_k, x_{k-d}, k) \end{array}\right]$, then (8) can be conveniently written as

$$\xi_1^{\mathrm{T}}(k) \left[\begin{array}{ccccc} -\delta_1^2 M_1^{\mathrm{T}} M_1 & \star & \star & \star & \star \\ -\delta_1^2 M_1^{\mathrm{T}} M_1 & -\delta_1^2 M_1^{\mathrm{T}} M_1 & \star & \star & \star \\ 0 & 0 & -\delta_{d1}^2 M_{d1}^{\mathrm{T}} M_{d1} & \star & \star \\ 0 & 0 & 0 & -\delta_{d1}^2 M_{d1}^{\mathrm{T}} M & \star \\ 0 & 0 & 0 & -\delta_{d1}^2 M_{d1}^{\mathrm{T}} M & I \end{array}\right] \xi_1(k) \le 0 \tag{9}$$

In addition,

$$\bar{f}^{\mathrm{T}}(\hat{x}_k, \hat{x}_{k-d}, k)\bar{f}(\hat{x}_k, \hat{x}_{k-d}, k) \le \delta_2^2 \hat{x}^{\mathrm{T}}(k) M_2^{\mathrm{T}} M_2 \hat{x}(k) + \delta_{d2}^2 \hat{x}^{\mathrm{T}}(k-d) M_{d2}^{\mathrm{T}} M_{d2} \hat{x}(k-d) \tag{10}$$

Let $\xi_2^{\mathrm{T}} = \left[\begin{array}{cccc} e^{\mathrm{T}}(k) & \hat{x}^{\mathrm{T}}(k) & e^{\mathrm{T}}(k-d) & \hat{x}^{\mathrm{T}}(k-d) & \bar{f}^{\mathrm{T}}(\hat{x}_k, \hat{x}_{k-d}, k) \end{array}\right]$, then (10) can be conveniently written as

$$\xi_2^T(k) \left[\begin{array}{ccccc} 0 & \star & \star & \star & \star \\ 0 & -\delta_2^2 M_2^{\mathrm{T}} M_2 & \star & \star & \star \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & -\delta_{d2}^2 M_{d2}^{\mathrm{T}} M_{d2} & \star \\ 0 & 0 & 0 & 0 & I \end{array}\right] \xi_2(k) \le 0 \tag{11}$$

where $\delta_1, \delta_2, \delta_{d1}, \delta_{d2}$ are the bounding parameter, and M_1, M_2, M_{d1}, M_{d2} are constant matrices such that $f(0,0,k)=0$ and $x=0$ is an equilibrium of system (2) for $d_k=0$.

The objective of this paper is to design observer-based non-fragile passive controller, substitute non-fragile observer (4) and controller (5) into system (2), and let $e(k)=x(k)-\hat{x}(k)$, then resulting error closed-loop system is obtain by

$$\begin{cases} e(k+1)=(G_0-LC_1+\Delta G_0-\Delta LC_1)e(k)+\Delta G_0\hat{x}(k)+(G_1+\Delta G_1)e(k-d)+ \\ \qquad\qquad \Delta G_1\hat{x}(k-d)+H_1\omega(k)+\bar{f}(x)-\bar{f}(\hat{x}) \\ \hat{x}(k+1)=(G_0+H_0K+H_0\Delta K)\hat{x}(k)+G_1\hat{x}(k-d)+(L+\Delta L)C_1e(k)+\bar{f}(\hat{x}(k)) \end{cases} \tag{12}$$

Before proceeding to this main results, the following useful assumption and lemmas are need.

Assumption 2.3. *Suppose that the matrix C_1 has full row rank (i.e. rank(C_1)=p). for convenience of discussion, the singular value decomposition of C_1 as follows:*

$$C_1=U\begin{bmatrix} S & 0 \end{bmatrix}V^{\mathrm{T}}$$

where is $S\in\mathbb{R}^{p\times p}$ a diagonal matrix with positive diagonal elements in decreasing order, $0\in\mathbb{R}^{p\times(n-p)}$ is a zero matrix, and $U\in\mathbb{R}^{p\times p}$ and $V\in\mathbb{R}^{n\times n}$ are unitary matrices.

Lemma 1. *[22] For a given $C_1\in\mathbb{R}^{p\times n}$ with rank(C_1)=p,assume that $X\in\mathbb{R}^{n\times n}$ is a symmetric matrix, then there exists a matrix $\hat{X}\in R^{p\times p}$ such that $C_1X=\hat{X}C_1$ if and only if*

$$X=V\begin{bmatrix} \hat{X}_{11} & 0 \\ 0 & \hat{X}_{22} \end{bmatrix}V^{\mathrm{T}}$$

where $\hat{x}_{11}\in\mathbb{R}^{p\times p}$ and $\hat{x}_{22}\in\mathbb{R}^{(n-p)\times(n-p)}$

Lemma 2. *[23] (Schur complement) For a given symmetric matrix $S=S^{\mathrm{T}}=\begin{bmatrix} S_{11} & S_{12} \\ S_{12}^{\mathrm{T}} & S_{22} \end{bmatrix}$ with $S_{11}\in\mathbb{R}^{r\times r}$, the following conditions are equivalent:*

(1) $S<0$
(2) $S_{11}<0, \quad S_{22}-S_{12}^{\mathrm{T}}S_{11}^{-1}S_{12}<0$
(3) $S_{22}<0, \quad S_{11}-S_{12}S_{22}^{-1}S_{12}^{\mathrm{T}}<0$

Lemma 3. *[24] For given matrices $\mathcal{Q}=\mathcal{Q}^{\mathrm{T}}$, H, and E, with appropriate dimensions*

$$\mathcal{Q}+HF(k)E+E^{\mathrm{T}}F^{\mathrm{T}}(k)H^{\mathrm{T}}<0$$

holds for all $F(k)$ satisfying $F^{\mathrm{T}}(k)F(k)\le I$ if and only if there exists $\varepsilon>0$

$$\mathcal{Q}+\varepsilon HH^{\mathrm{T}}+\varepsilon^{-1}E^{\mathrm{T}}E<0$$

Definition 4. The systems (2) is called passive if there exists a scalar $\beta\ge 0$ such that

$$\sum_{k=0}^{\infty}\omega(k)z(k)\ge\beta\,,\ \forall\omega\in\mathbb{I}_2[0,\,\infty]$$

where β is some constant which depends on the initial condition of the system.

3 Main Results

Theorem 5. *For system(2) and observer (3), if there exist two symmetric and positive matrices* $\bar{R} \in \mathbb{R}^{n\times n}$, $\bar{P} \in \mathbb{R}^{n\times n}$, *two real matrices* $Y_1 \in \mathbb{R}^{m\times n}$, $Y_2 \in \mathbb{R}^{n\times p}$ *and three positive constants* ε_1, ε_2, *and* ε_3 *such that the following holds:*

$$\Xi = \begin{bmatrix} \Xi_{11} & \Xi_{21}^{\mathrm{T}} \\ \Xi_{21} & \Xi_{22} \end{bmatrix} < 0 \tag{13}$$

then there exist two gains $K = Y_1\bar{P}^{-1}$, *and* $L = Y_2 U S \hat{X}_{11}^{-1} S^{-1} U^{\mathrm{T}}$, *such that system is asymptotically passive stable.*
where

$$\Xi_{11} = \begin{bmatrix} W_1-\bar{R} & \star & \star & \star & \star & \star & \star & \star \\ 0 & W_2-\bar{P} & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star & \star \\ -C_2\bar{R} & -C_2\bar{P} & 0 & 0 & -H_2-H_2^{\mathrm{T}} & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star \\ G_0\bar{R}-Y_2C_1 & 0 & G_0 & 0 & H_1 & I & -I & -\bar{R} \end{bmatrix}$$

$$\Xi_{21} = \begin{bmatrix} Y_2C_1 & G_0\bar{P}+H_0Y_1 & 0 & G_1 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_1 M_0^{\mathrm{T}} \\ E_0\bar{R}-E_2C_1\bar{R} & E_0\bar{P} & E_1 & E_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_2C\bar{R} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E_3\bar{P} & 0 & 0 & 0 & 0 & 0 & 0 \\ \delta_1 M_1\bar{R} & 0 & \delta_{d1}M_{d1} & \delta_{d1}M_{d1}+\delta_{d2}M_{d2} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Xi_{22} = \begin{bmatrix} -\bar{P} & \star & \star & \star & \star & \star & \star & \star \\ 0 & -\varepsilon_1 I & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -\varepsilon_1 I & \star & \star & \star & \star & \star \\ \varepsilon_1 M_0^T & 0 & 0 & -\varepsilon_1 I & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & -\varepsilon_1 I & \star & \star & \star \\ \varepsilon_1 (H_0M_0)^T & 0 & 0 & 0 & 0 & -\varepsilon_1 I & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_1 I & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_1 I \end{bmatrix}$$

Proof: Choose a Lyapunov function candidate for the system (12) as follow:

$$
\begin{aligned}
V &= V_1 + V_2 + V_3 + V_4 \\
V_1 &= e^{\mathrm{T}}(k) R e(k) \\
V_2 &= \sum_{i=k-h}^{k-1} e^{\mathrm{T}}(i) Q_1 e(i) \\
V_3 &= \hat{x}^{\mathrm{T}}(k) P \hat{x}(k)
\end{aligned}
$$

$$
V_4 = \sum_{i=k-h}^{k-1} x^{\mathrm{T}}(i) Q_2 x(i)
$$

where $R = R^{\mathrm{T}} > 0$, $Q_1 = Q_1^{\mathrm{T}} > 0$, $Q_2 = Q_2^{\mathrm{T}} > 0$, and $P = P^{\mathrm{T}} > 0$.
Define vector

$$
\xi(k) = \begin{bmatrix} e(k) \\ \hat{x}(k) \\ e(k-d) \\ \hat{x}(k-d) \\ \omega(k) \\ \bar{f}(x_k, x_{k-d}, k) \\ \bar{f}(\hat{x}_k, \hat{x}_{k-d}, k) \end{bmatrix}, \quad \theta_1 = \begin{bmatrix} (G_0 - LC_1 + \Delta G_0 - \Delta L C_1)^{\mathrm{T}} \\ \Delta G_0^{\mathrm{T}} \\ (G_1 + \Delta G_1)^{\mathrm{T}} \\ \Delta G_1^{\mathrm{T}} \\ H_1^{\mathrm{T}} \\ I \\ -I \end{bmatrix}^{\mathrm{T}}
$$

$$
\theta_2 = \begin{bmatrix} H(L + \Delta L)C & G_0 - H_0 K & 0 & G_1 & 0 & 0 & I \end{bmatrix}
$$

Therefore,

$$
\begin{aligned}
\Delta V &= \Delta V_1 + \Delta V_2 + \Delta V_3 + \Delta V_4 \\
&= \xi^{\mathrm{T}}(k) \theta_1^{\mathrm{T}} R \theta_1 \xi(k) + e^{\mathrm{T}}(k) Q_1 e(k) - e^{\mathrm{T}}(k-d) Q_1 e(k-d) + \\
&\quad \xi^{\mathrm{T}}(k) \theta_2^{\mathrm{T}} P \theta_2 \xi(k) + \hat{x}^{\mathrm{T}}(k) Q_2 \hat{x}(k) - \hat{x}^{\mathrm{T}}(k-d) Q_2 \hat{x}(k-d) \\
&= \xi^{\mathrm{T}}(k) \Pi_1 \xi(k)
\end{aligned} \tag{14}
$$

On one hand, the sufficient condition of stability $\Delta V < 0$, implies that $\Pi_1 < 0$, that is

$$
\Pi_1 := \begin{bmatrix}
Q_1 - R & \star & \star & \star & \star & \star & \star & \star & \star \\
0 & Q_2 - P & \star & \star & \star & \star & \star & \star & \star \\
0 & 0 & -Q_1 & \star & \star & \star & \star & \star & \star \\
0 & 0 & 0 & -Q_2 & \star & \star & \star & \star & \star \\
0 & 0 & 0 & 0 & 0 & \star & \star & \star & \star \\
0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\
G_0 - LC_1 + \Delta G_0 - \Delta L C_1 & \Delta G_0 & G_1 + \Delta G_1 & \Delta G_1 & H_1 & I & -I & -R^{-1} & \star \\
(L + \Delta L)C & G_0 - H_0 K & 0 & G_1 & 0 & 0 & I & 0 & -P^{-1}
\end{bmatrix} < 0 \tag{15}
$$

On the other hand, utilizing (14) with $\omega(k) \in \mathbb{l}_2[0, +\infty] \neq 0$, one is obtained by

$$\begin{aligned}
&\Delta V - 2z^T(k)\omega(k) \leq \\
&\xi^{\mathrm{T}}(k)\left\{\begin{bmatrix} Q_1 - R & \star & \star & \star & \star & \star & \star \\ 0 & Q_2 - P & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \right. \\
&\left. \theta_1^{\mathrm{T}} R\theta_1 + \theta_2^{\mathrm{T}} P\theta_2 - 2\theta_3^{\mathrm{T}}\theta_4 \right\} \xi(k) = \xi^{\mathrm{T}}(k)\Pi_2\xi(k)
\end{aligned} \tag{16}$$

If $\Pi_2 < 0$, then $\Delta V(k) - 2z^{\mathrm{T}}(k)\omega(k) < 0$ and from which it follows that

$$\sum_{k=0}^{\infty} \omega(k)z(k) > 1/2 \sum_{k=0}^{\infty} \Delta V = 1/2[V(0) - V(\infty)] \tag{17}$$

Due to $V(k) > 0$ for $x \neq 0$ and $V(k) = 0$ for $x = 0$, if follows as $k \to \infty$ that system (12) is strictly passive. In virtue of Definition 4, the strictly passive condition is guaranteed if $\Pi_2 < 0$ and it can be expressed conveniently as

$$\begin{bmatrix} Q_1 - R & \star & \star & \star & \star & \star & \star \\ 0 & Q_2 - P & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \theta_1^{\mathrm{T}} R\theta_1 + \theta_2^{\mathrm{T}} P\theta_2 - 2\theta_3^{\mathrm{T}}\theta_4 < 0 \tag{18}$$

Application of the Lemma 2 to (18) puts it into the form:

$$\begin{bmatrix} Q_1 - R & \star & \star & \star & \star & \star & \star & \star & \star \\ 0 & Q_2 - P & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star & \star & \star \\ -C_2 & -C_2 & 0 & 0 & -H_2 - H_2^{\mathrm{T}} & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ G_0 - LC_1 + \Delta G_0 - \Delta LC_1 & \Delta G_0 & G_1 + \Delta G_1 & \Delta G_1 & H_1 & I & -I & -R^{-1} & \star \\ (L + \Delta L)C_1 & G_0 - H_0(K + \Delta K) & 0 & G_1 & 0 & 0 & I & 0 & -P^{-1} \end{bmatrix} < 0 \tag{19}$$

Substituting the uncertainty structure into (19) and rearranging, we get

$$\begin{bmatrix} Q_1-R & \star & \star & \star & \star & \star & \star & \star & \star \\ 0 & Q_2-P & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star & \star & \star \\ -C_2 & -C_2 & 0 & 0 & -H_2 & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ G_0-LC_1 & 0 & G_0 & 0 & H_1 & I & -I & -R^{-1} & \star \\ LC & G_0-H_0K & 0 & G_1 & 0 & 0 & I & 0 & -P^{-1} \end{bmatrix} \tag{20}$$
$$+\theta_5 F(k)\theta_6+\theta_6^{\mathrm{T}}F^{\mathrm{T}}(k)\theta_5^{\mathrm{T}}+\theta_7 F(k)\theta_8+\theta_8^{\mathrm{T}}F^{\mathrm{T}}(k)\theta_7^{\mathrm{T}}+\theta_9 F(k)\theta_{10}+$$
$$\theta_{10}^{\mathrm{T}}F^{\mathrm{T}}(k)\theta_9^{\mathrm{T}}<0$$

where

$$\theta_5=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & M_0^{\mathrm{T}} & 0 \end{bmatrix}^{\mathrm{T}}$$
$$\theta_6=\begin{bmatrix} E_0-E_2C_1 & E_0 & E_1 & E_1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\theta_7=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & M_0^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$
$$\theta_8=\begin{bmatrix} E_2C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\theta_9=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (H_0M_0)^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$
$$\theta_{10}=\begin{bmatrix} 0 & E_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Then by Lemma 3, the inequality (20) holds if and only if for some $\varepsilon_i>0$ (i=1,$\cdots$, 3)

$$\Pi_3+\varepsilon_1\theta_5\theta_5^{\mathrm{T}}+\varepsilon_1^{-1}\theta_6^{\mathrm{T}}\theta_6+\varepsilon_2\theta_7\theta_7^{\mathrm{T}}+\varepsilon_2^{-1}\theta_8^{\mathrm{T}}\theta_8+\varepsilon_3\theta_9\theta_9^{\mathrm{T}}+\varepsilon_3^{-1}\theta_{10}^{\mathrm{T}}\theta_{10}<0 \tag{21}$$

where

$$\Pi_3:=\begin{bmatrix} Q_1-R & \star & \star & \star & \star & \star & \star & \star & \star \\ 0 & Q_2-P & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star & \star & \star \\ -C_2 & -C_2 & 0 & 0 & -H_2 & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ G_0-LC_1 & 0 & G_0 & 0 & H_1 & I & -I & -R^{-1} & \star \\ LC_1 & G_0-H_0K & 0 & G_1 & 0 & 0 & I & 0 & -P^{-1} \end{bmatrix}$$

On using the Lemma 2, it becomes that

$$\begin{bmatrix} \Pi_3 & \star & \star & \star & \star & \star & \star \\ \varepsilon_1\theta_5^{\mathrm{T}} & -\varepsilon_1 I & \star & \star & \star & \star & \star \\ \theta_6 & 0 & -\varepsilon_1 I & \star & \star & \star & \star \\ \varepsilon_2\theta_7^{\mathrm{T}} & 0 & 0 & -\varepsilon_2 I & \star & \star & \star \\ \theta_8 & 0 & 0 & 0 & -\varepsilon_2 I & \star & \star \\ \varepsilon_3\theta_9^{\mathrm{T}} & 0 & 0 & 0 & 0 & -\varepsilon_3 I & \star \\ \theta_{10} & 0 & 0 & 0 & 0 & 0 & -\varepsilon_3 I \end{bmatrix} < 0 \tag{22}$$

Thirdly, introduce nonlinearities (8) and (10) into (22), then

$$\begin{bmatrix} \Pi_4 & \star & \star & \star & \star & \star & \star \\ \varepsilon_1\theta_5^{\mathrm{T}} & -\varepsilon_1 I & \star & \star & \star & \star & \star \\ \theta_6 & 0 & -\varepsilon_1 I & \star & \star & \star & \star \\ \varepsilon_2\theta_7^{\mathrm{T}} & 0 & 0 & -\varepsilon_2 I & \star & \star & \star \\ \theta_8 & 0 & 0 & 0 & -\varepsilon_2 I & \star & \star \\ \varepsilon_3\theta_9^{\mathrm{T}} & 0 & 0 & 0 & 0 & -\varepsilon_3 I & \star \\ \theta_{10} & 0 & 0 & 0 & 0 & 0 & -\varepsilon_3 I \end{bmatrix} < 0 \tag{23}$$

where

$$\Pi_4 := \begin{bmatrix} \varphi_1 & \star & \star & \star & \star & \star & \star & \star & \star \\ -\delta_1^2 M_1^{\mathrm{T}} M_1 & \varphi_2 & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & \varphi_3 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -\delta_{d1}^2 M_{d1}^{\mathrm{T}} M_{d1} & \varphi_4 & \star & \star & \star & \star & \star \\ -C_2 & -C_2 & 0 & 0 & -H_2 & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ G_0 - LC_1 & \Delta G_0 & G_0 & 0 & H_1 & I & -I & -R^{-1} & \star \\ LC_1 & G_0 - H_0 K & 0 & G_1 & 0 & 0 & I & 0 & -P^{-1} \end{bmatrix} < 0$$

$$\begin{aligned} \varphi_1 &= Q_1 - R - \delta_1^2 M_1^{\mathrm{T}} M_1 \\ \varphi_2 &= Q_2 - P - \delta_1^2 M_1^{\mathrm{T}} M_1 - \delta_2^2 M_2^{\mathrm{T}} M_2 \\ \varphi_3 &= -Q_1 - \delta_{d1}^2 M_{d1}^{\mathrm{T}} M_{d1} \\ \varphi_4 &= -Q_2 - \delta_{d1}^2 M_{d1}^{\mathrm{T}} M_{d1} - \delta_{d2}^2 M_{d2}^{\mathrm{T}} M_{d2} \end{aligned}$$

Pre-multiplying and post-multiplying (23) by $\mathrm{diag}\left\{ \begin{array}{cc} R^{-1} & P^{-1} \end{array} \overbrace{\begin{array}{ccc} I & \cdots & I \end{array}}^{13\ columns} \right\}$, and from Lemma 1, the condition $C_1\bar{R} = \hat{\bar{R}}C_1$ holds, moreover, setting $Y_2 = L\hat{\bar{R}}$. In the meanwhile, define $\bar{P} = P^{-1}, \bar{R} = R^{-1}, Y_1 = K\bar{P}, W_1 = R^{-1}Q_1R^{-1}, W_2 = P^{-1}Q_2P^{-1}$, it is seen that (23)<0 is equivalent to (13), the means that the system (12) is asymptotically passive stable.

Theorem 6. *For system(2) and observer (3), if there exist two symmetric and positive matrices* $\bar{R} \in \mathbb{R}^{n\times n}$, $\bar{P} \in \mathbb{R}^{n\times n}$, *two real matrices* $Y_3 \in \mathbb{R}^{m\times n}$, $Y_4 \in \mathbb{R}^{n\times p}$ *and three positive constants* ε_4,

ε_5, and ε_6 such that the following holds:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{21}^{\mathrm{T}} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} < 0 \tag{24}$$

then there exist two gains $K = Y_3\bar{P}^{-1}$, and $L = Y_4 U S \hat{X}_{11}^{-1} S^{-1} U^{\mathrm{T}}$, such that system is asymptotically passive stable.
where

$$\Omega_{11} = \begin{bmatrix} W_3 - \bar{R} & \star & \star & \star & \star & \star & \star & \star \\ 0 & W_4 - \bar{P} & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -Q_1 & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & -Q_2 & \star & \star & \star & \star \\ -C_2\bar{R} & -C_2\bar{P} & 0 & 0 & -H_2 - H_2^{\mathrm{T}} & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star \\ G_0\bar{R} - Y_4C_1 & 0 & G_0 & 0 & H_1 & I & -I & -\bar{R} \end{bmatrix}$$

$$\Omega_{21} = \begin{bmatrix} Y_4C_1 & G_0\bar{P} + H_0Y_3 & 0 & G_1 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_4 M_0^{\mathrm{T}} \\ E_0\bar{R} - E_2C_1\bar{R} & E_0\bar{P} & E_1 & E_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_2C\bar{R} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E_4Y_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ \delta_1 M_1\bar{R} & 0 & \delta_{d1}M_{d1} & \begin{matrix}\delta_{d1}M_{d1}+\\ \delta_{d2}M_{d2}\end{matrix} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Omega_{22} = \begin{bmatrix} -\bar{P} & \star & \star & \star & \star & \star & \star & \star \\ 0 & -\varepsilon_4 I & \star & \star & \star & \star & \star & \star \\ 0 & 0 & -\varepsilon_4 I & \star & \star & \star & \star & \star \\ \varepsilon_4 M_0^T & 0 & 0 & -\varepsilon_5 I & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & -\varepsilon_5 I & \star & \star & \star \\ \varepsilon_4 (H_0 M_0)^T & 0 & 0 & 0 & 0 & -\varepsilon_6 I & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_6 I & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{bmatrix}$$

Proof: Theorem 6 's proof is same to Theorem 5, so is omitted.

4 Numerical Example

Consider the plant of sampled-data system (1) with the parameters:

$$A_0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 1 \end{bmatrix}, \; A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & -3 & 1 \end{bmatrix}, \; B_0 = \begin{bmatrix} 0.1 & 0.1 \\ 0 & 1.2 \\ 0.5 & 0.1 \end{bmatrix}, \; B_1 = \begin{bmatrix} 0.1 & 0.2 \\ 0 & 1.5 \\ 0 & 0.1 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 1.5 & 0.6 & 1.0 \\ 0.5 & 0.6 & 0 \end{bmatrix}, \; C_2 = \begin{bmatrix} 3 & 4 & 1 \end{bmatrix}, \; H_2 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}, \; H_3 = 0.5$$

By Theorem 1 has a solution:

$$\bar{P} = \begin{bmatrix} 3.8697 & -2.7891 & -0.4465 \\ -2.7891 & 2.0973 & -0.0138 \\ -0.4465 & -0.0138 & 1.3967 \end{bmatrix}, \; Y_1 = \begin{bmatrix} -13.8330 & 9.8716 & 2.0314 \\ 1.4006 & -0.8149 & -0.9412 \end{bmatrix}$$

$$\bar{R} = \begin{bmatrix} 0.3864 & -0.3364 & -0.3658 \\ -0.3364 & 0.3753 & 0.2113 \\ -0.3658 & 0.2113 & 0.4907 \end{bmatrix}, \; \hat{X}_{11} = \begin{bmatrix} 0.0059 & 0.0333 \\ 0.0333 & 0.2139 \end{bmatrix}, \; \hat{X}_{22} = 1.0326$$

The non-fragile passive observer-based control are given by

$$L = Y_2 U S \hat{X}_{11}^{-1} S^{-1} U^{\mathrm{T}} = \begin{bmatrix} -6.5298 & -9.0418 \\ 2.7828 & 3.9663 \\ 8.6606 & 11.5095 \end{bmatrix}, \; K = Y_1 \bar{P}^{-1} = \begin{bmatrix} -1.3147 & 2.9654 & 1.0635 \\ 0.3124 & 0.0231 & -0.5738 \end{bmatrix}$$

We present design method of observer-based non-fragile passive controller in this paper, the simulation results are given in Figure 1. From Figure 1, it can be seen the state estimation $\hat{x}(t)$ has a good trace performance with the external disturbance and input nonlinearity.

5 Application to Stabilization of An Inverted Pendulum on A Cart

An inverted pendulum on a cart [25] is depicted in Figure 2

In this model, a pendulum is conjuncted to the topside of a cart by a pivot, which is allowed to swing in the $xy-$plane. A force u acts on the cart in the x direction, in order to keep the pendulum balance upright. $x(t)$ is the displacement between central mass of cart and the origin 0; θ is the angle of the pendulum from the top vertical.

Which is described by the following dynamics by applying Newtons Second Law

$$(M+m)\,\ddot{x} + ml\ddot{\theta}\cos\theta - ml\dot{\theta}^2\sin\theta = u$$

$$ml\ddot{x}\cos\theta + \frac{4}{3}ml^2\ddot{\theta} - mgl\sin\theta = 0$$

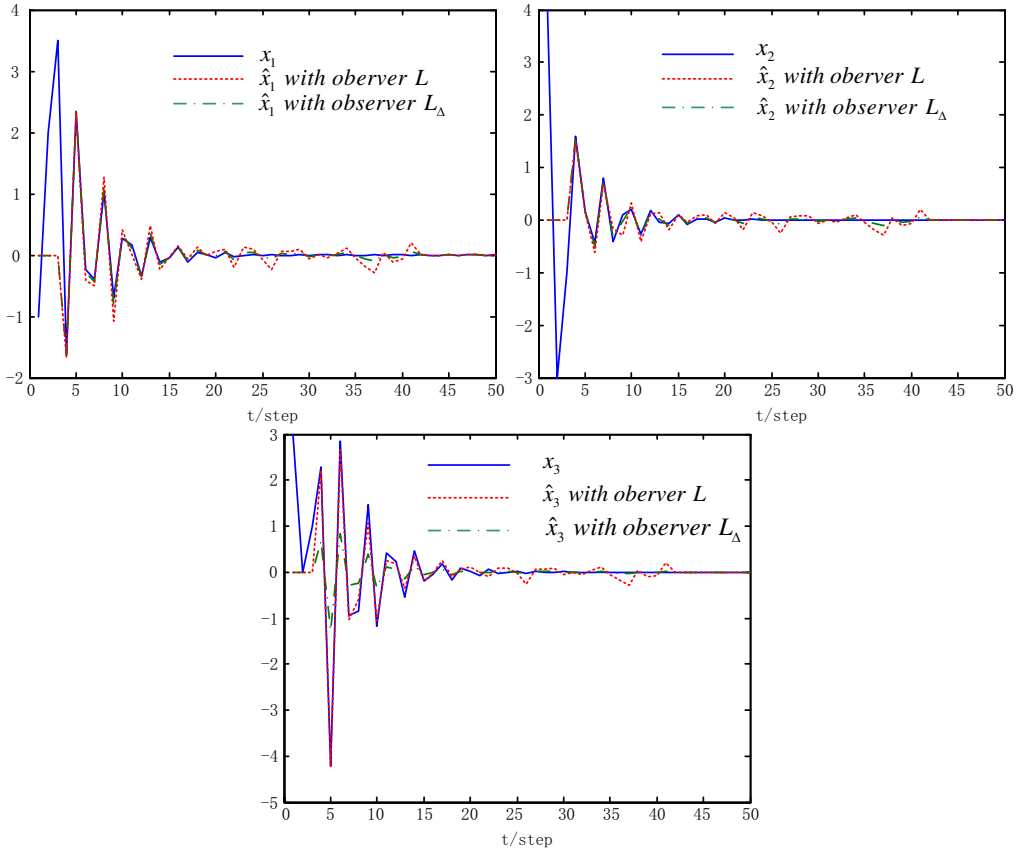


Figure 1: The simulation of non-fragile observer with additional perturbation

Now, by selecting state variables $z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \theta & \dot{\theta} \end{bmatrix}^{\mathrm{T}}$ and by linearizing the above model at the equilibrium point $z = 0$, we obtain the following state-space model:

$$\dot{z}(t) = \begin{bmatrix} 0 & 1 \\ \frac{3(M+m)g}{l(4M+m)} & 0 \end{bmatrix} z(t) + \begin{bmatrix} 0 \\ -\frac{3}{l(4M+m)} \end{bmatrix} u(t) \tag{25}$$

Here the parameters are selected in Table 1, by assuming the sampling time to be $T_s = 0.1$ s, the discretized model for the above pendulum system in (21) is given by

$$x(k+1) = \begin{bmatrix} 1.0877 & 0.1029 \\ 1.7794 & 1.0877 \end{bmatrix} x(k) + \begin{bmatrix} -0.0000 \\ -0.0182 \end{bmatrix} u(k) \tag{26}$$

The poles of the system are 0.6598 and 1.5156, thus this discretized system is unstable. It is assumed that a non-fragile control law with additive form is given by

$$u(k) = \begin{bmatrix} 219.023 & 49.786 \end{bmatrix} x(k) \tag{27}$$

The other non-fragile control law with multiplicative form is given by

$$u(k) = \begin{bmatrix} 199.755 & 52.011 \end{bmatrix} x(k) \tag{28}$$

Table 1. An inverted pendulum parameters

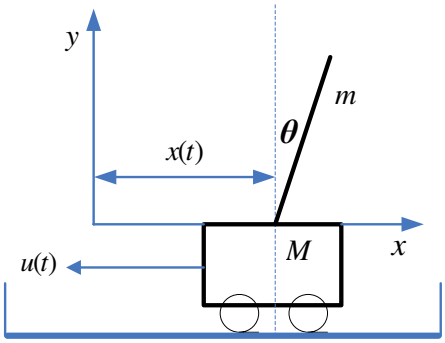
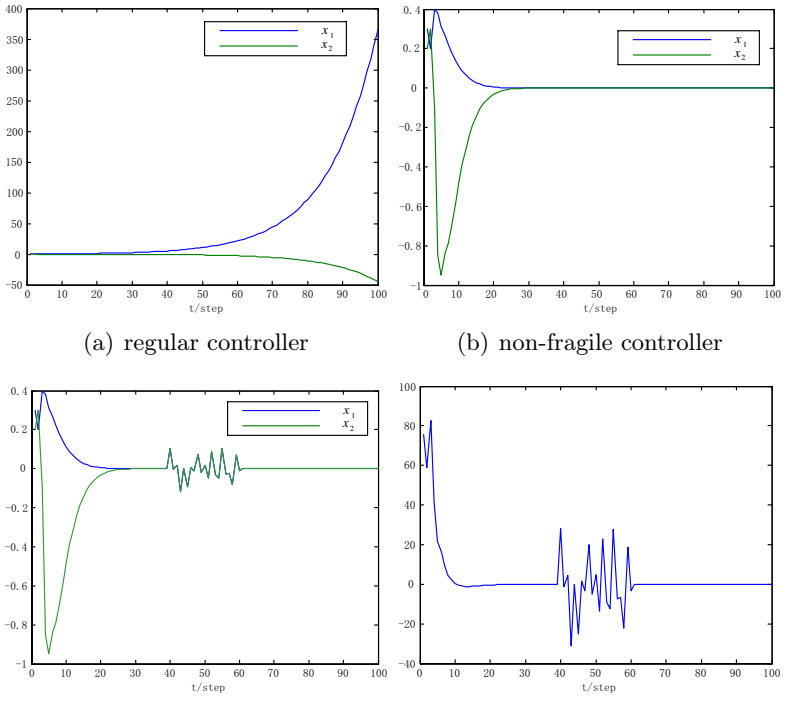


Figure 2: A pendulum system perturbation

System parameter	Values
Mass of the cart M (kg)	8.0
Mass of the pendulum (kg)	2.0
Half length of the pendulum (m)	0.5
Acceleration of gravity (m/s^2)	9.8



(a) regular controller (b) non-fragile controller

(c) non-fragile controller with non-liearity (d) control input

Figure 3: State response of the pendulum system and control input

The simulation results are given in Figure 3. In fact, for sampling period $T_s = 0.1s$, LMI (13) remain solvable. State response curve of regular controller is divergent in Figure 3(a), however, the curve is convergent for non-fragile controller in Figure 3(b). Furthermore, it is still convergent, when there exists a nonlinear perturbation in Figure 3(c). A corresponding control input is shown in Figure 3(d).

6 Conclusions

The problem of observer-based non-fragile passive control of uncertain nonlinear sampled-data system with time-delay has been studied. A LMI based approach to designing state observer and non-fragile controller, which ensure the passivity of the resulting error closed-loop system has been developed. A numerical example has been provided to demonstrate the effectiveness and applicability of the proposed approach.

Acknowledgment

This work is partially supported by National Defense 973 Basic Research Development Program of China (6131380101), National Natural Science Foundation of China (61203121,51307045),and the authors are indebted to the editor and reviewers for their valuable comments and suggestions.

Bibliography

[1] Chen, T.; Francis, B. A. (1995); Optimal Sampled-Data Control System, *Springer*, New York.

[2] Wu, J. F.; Wang, Q.; Chen, S. B. (2001); Robust Stability for Sampled-Data Systems, *Control Theory & Applications*, ISSN 1000-8152, 18(8):99-102.

[3] Wu, J. F.; Wang, Q.; Chen, S. B. (2002); Robust control of a class of sampled-data systems with structured uncertainty, *Control and Decision*, ISSN 1001-0920, 17(5):681-685.

[4] Liu, F. C.; Yao, Y.; He, F. H. (2009); Robust $\mathcal{H}_\infty$ controller design for sampled-data systems with parametric uncertainties, *Journal of Systems Engineering and Electronics*, ISSN 1004-4132, 20(2):371-378.

[5] Wang, G. X.; Liu, Y. W.; Zhen, H. (2006); $\mathcal{H}_\infty$ mixed sensitivity design for sampled-data systems, *Control Theory & Applications*, ISSN 1000-8152, 23(3):351-354.

[6] Wang, S. G. (2011); Non-fragile $\mathcal{H}_\infty$ control with pole constraints for a class of nonlinear sampled-data system, *Lecture Notes in Electrical Engineering*, ISSN 1876-1100, 87(2):587-594.

[7] Wang, S. G.; Bi, Y. L.; Li, Y. S. (2014); Improvements on Robust Stability of Sampled-Data System with Long Time Delay, *Mathematical Problems in Engineering*, ISSN 1024-123X, 2014, Article ID 580768, 7 pages, http://dx.doi.org/10.1155/2014/580768.

[8] Abid, H.; Chtourou, M.; Toumi, A. (2008); Robust Fuzzy Sliding Mode Controller for Discrete Nonlinear Systems, *International Journal of Computers, Communications & Control*, ISSN 1841-9836, III(1):6-20.

[9] Xianga, Z. (2010); Robust Control of Particle Size Distribution in Aerosol Processes, *International Journal of Computers, Communications & Control*, ISSN 1841-9836, V(3):385-397.

[10] Ngo, T.; Wang, Y.; Mai, T. L.; Nguyen, M. H.; Chen, J. (2012); Robust Adaptive Neural-Fuzzy Network Tracking Control for Robot Manipulator, *International Journal of Computers, Communications & Control*, ISSN 1841-9836, 7(2):341-352.

[11] Farokhi, M. H.; Vasegh, N. (2014); Robust PID Stabilization of Linear Neutral Time-delay Systems, *International Journal of Computers, Communications & Control*, ISSN 1841-9836, 9(2):201-208.

[12] Keel, L. H.; Bhattacharyya, S. P. (1997); Robust, fragile, or optimal, *IEEE Transactions Automatic Control*, ISSN 0018-9286, 42(8):2678-2683.

[13] Yang, G. H.; Wang, J. L.; Lin, C. (2000); $\mathcal{H}_\infty$ control for linear Systems with additive gain variations, *International Journal of Control*, ISSN 0020-7179, 73(16):1500-1506.

[14] Yang, G. H.; Wang, J. L. (2001); Non-fragile control for linear Systems with multiplicative controller gain variations, *Automatica*, ISSN 0005-1098, 37(5):727-737.

[15] Wang, S. G.; Wu, J. F. (2011); Non-fragile $\mathcal{H}_\infty$ filtering for a class of sampled-data system with long time-delay, *ICIC Express Letters Part B: Appliations*, ISSN 2185-2766, 2(6):1447-1452.

[16] Wang, S. G.; Li, Y. S.; Wu, Y. (2013); Non-fragile Passive Filtering for Sampled-data System with Long Time-delay Subject to Nonlinearity, *Recent Patents on Electrical & Electronic Engineering*, ISSN 1874-4761, 6(3):196-202.

[17] Chen, J. D.; Yang, C. D.; Lien, C. H.; Horng, J. H. (2008); New delay-dependent non-fragile $\mathcal{H}_\infty$ observer-based control for continous time-delay systems, *Information Sciences*, ISSN 1968-2013, 178(24):4799-4706.

[18] Ibrir, S.; Yang, D. S. (2008); Novel LMI conditions for observer-based stabilization of Lipschitzian nonlinear systems and uncertain linear systems in discrete-time, *Applied Mathematics and Computation*, ISSN 1975-2013, 206(2):579-588.

[19] Wang, S. G.; Wu, J. F. (2011); Observer-based Non-fragile $\mathcal{H}_\infty$ control for a class of uncertain time-delay sampled-data systems, *Systems Engineering and Electronics*, ISSN 1001-506X, 33(6):1358-1361.

[20] Shih, K. S.; Li, T. S.; Tsai S. H. (2012); Observer-Based Adaptive Fuzzy Robust Controller with Self-Adjusted Membership Functions for a Class of Uncertain MIMO Nonlinear Systems: a PSO-SA Method, *International Journal of Innovative Computing Information and Control*, ISSN 1349-4198, 8(2):1025-1036.

[21] Wang, S. G.; Bi, Y. L. (2012); Robust Output Feedback Control via Pole Constraints for Sampled-Data System with Long Time-Delay, *Journal of Computational and Theoretical Nanoscience*, ISSN 1546-1955, 10(12):2926-2930.

[22] Gu, D. W.; Poon, F. W. (2003); Robust stabilization for a class of discrete-time non-linear systems via output feedback: the unified LMI approach, *International Journal of Control*, ISSN 0020-7179, 76(2):105-115.

[23] Albert. (1969); Conditions for positive and non-negative definiteness in terms of pseudoinverses, *International Journal of Control*, ISSN 1095-712X, 17(2):434-440.

[24] Barmish, B. R. (1985); Necessary and sufficient conditions for quadratic stabilizability of an uncertain system, *Journal of Optimization Theory and Applications*, ISSN 0022-3239, 46(4):399-408.

[25] Gao, H. J.; Chen, T. W. (2007); New results on stability of discrete-time systems with time-varying state delay, *IEEE Transactions on Automatic Control*, ISSN 0018-9286, 52(2):328-334.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):497-509, August, 2014.

Optimized Branch and Bound for Path-wise Test Data Generation

Y.W. Wang, Y. Xing, Y.Z. Gong, X.Z. Zhang

Ya-Wen Wang (1,2), **Ying Xing*** (1,3), **Yun-Zhan Gong** (1), **Xu-Zhou Zhang** (1)
(1) State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications, Beijing, China
10 Xitucheng Road, Beijing, China
E-mail: wangyawen@bupt.edu.cn, gongyz@bupt.edu.cn, laomao22311@126.com
(2) State Key Laboratory of Computer Architecture
Institute of Computing Technology, Chinese Academy of Sciences
(3) School of Electronic and Information Engineering
Liaoning Technical University
*Corresponding author: lovelyjamie@yeah.net

Abstract: The increasing complexity of real-world programs necessitates the automation of software testing. As a basic problem in software testing, the automation of path-wise test data generation is especially important, which is in essence a constraint satisfaction problem solved by search strategies. In this paper, the search algorithm branch and bound is introduced and optimized to tackle the problem of path-wise test data generation. The optimized branching operation is fulfilled by a dynamic variable ordering algorithm with a heuristic rule to break ties. The optimized bounding operation is accomplished by analyzing the results of interval arithmetic. In order to facilitate the search methods, the solution space is represented as state space. Experimental results prove the effectiveness of the optimized branching and bounding operations, and show that the proposed method outperformed some other methods used in test data generation. The results also demonstrate that the proposed method is applicable in engineering.
Keywords: test data generation, constraint satisfaction problem, branch and bound, state space search.

1 Introduction

With the surge of increasingly complex real-world software, software testing plays a more and more important role in the process of software development [1]. In 2002, National Institute of Standards and Technology (NIST) found that over one third of the cost of software failure could be eliminated by an improved testing infrastructure [2]. But manual testing is time-consuming and error-prone, and is even impracticable for real-world programs. So the automation of testing is of crucial concern [3]. Furthermore, as a basic problem in software testing, path-wise test data generation (denoted as Q) is of particular importance because many problems in software testing can be transformed into Q.

The methods of solving Q can be categorized as dynamic and static. The dynamic methods require the actual execution of the program under test (PUT), and the metaheuristic (MHS) [4] methods are very popular. Recently, the MHS method particle swarm optimization (PSO) [5] has become a hot research topic due to its convenient implementation and faster convergence speed. But dynamic methods often consume a large number of iterations, and the definition of objective function is also a big problem. The static methods utilize techniques including symbolic execution [6] and interval arithmetic [7] to analyze the PUT without executing it. The process of generating test data is definite with relatively less cost. They abstract the constraints to be satisfied, and propagate and solve these constraints to obtain the test data. Due to their precision in generating test data and the ability to prove that some paths are infeasible, the static methods

have been widely studied by many researchers. Demillo and Offutt [8] proposed a fault-based technique that used algebraic constraints to describe test data designed to find particular types of faults. Gotlieb et al. [9] introduced static single assignment into a constraint system and solved the system. Cadar et al. from Stanford University proposed a symbolic execution tool named KLEE [10] and employed a variety of constraint solving optimizations. In 2013, Wang et al. [11] proposed an interval analysis algorithm using forward dataflow analysis. No matter what techniques are adopted, static methods require a strong constraint solver.

In this paper, considering the drawbacks of the dynamic methods and the demand for static methods, we propose a new static test data generation method based on Code Test System (CTS)(http://ctstesting.com), which is a practical tool to test codes written in C programming language. Our contribution is threefold. First, path-wise test data generation is defined as a constraint satisfaction problem (CSP). Two techniques (state space search and branch and bound) in artificial intelligence are integrated to tackle the CSP. Second, the branching operation is optimized with a heuristic variable ordering algorithm. Third, the bounding operation is optimized in different stages of the search to reduce the search space greatly. Through experimental results, we try to evaluate the performance of our method, especially the optimized branching and bounding operations. We also make comparison experiments to find whether our method outperforms some currently existing test data generation methods in terms of coverage.

2 Problem definition and solving strategies

2.1 Problem definition

A control flow graph (CFG) for a program P is a directed graph G=(N,E,i,o), where N is a set of nodes, E is a set of edges, and i and o are respective unique entry and exit nodes to the graph. Each node $n\in N$ is a statement in the program, with each edge $e=(n_r,\ n_t)\in E$ representing a transfer of control from node n_r to node n_t. Nodes corresponding to decision statements such as *if* statements are branching nodes. Outgoing edges from these nodes are referred to as branches. A path through a CFG is a sequence $p=(n_1,\ n_2,\ \ldots,\ n_q)$, such that for all r, $1\leq r<q$, $(\mathrm{n}_r,\ \mathrm{n}_{r+1})\in E$. A path p is regarded as feasible if there exists a program input for which p is traversed, otherwise p is regarded as infeasible. The problem Q is in essence a CSP [12]. $X$is a set of variables $\{x_1,\ x_2,\ldots,\ x_n\}$, $D=\{D_1,\ D_2,\ldots,\ D_n\}$ is a set of domains, and $D_i\in D$ (i=1,2,...,n) is a finite set of possible values for x_i. For each path, D is defined based on the variables' acceptable ranges. One solution to the problem is a set of values to instantiate each variable inside its domain denoted as$\{x_1\mapsto V_1,x_2\mapsto V_2,\ldots,x_n\mapsto V_n\}$.

A CSP is generally solved by search strategies, among which backtracking algorithms are widely used. In this paper, state space search [13] and the backtracking algorithm branch and bound (BB) [14] are introduced to solve the CSP. The process of exploring the solution space is represented as state space search, as introduced in our previous work [15] . This representation will facilitate the implementation of BB. In classical BB search, nodes are always fully expanded, that is, for a given leaf node, all child nodes are immediately added to the so called open list. However, considering that one solution is enough for path-wise test data generation, best-first-search is our first choice. To find the best, ordering of variables is required for branching to prune the branches stretching out from unneeded variables. In addition, as the domain of a variable is a finite set of possible values which may be quite large, bounding is necessary to cut the unneeded or infeasible solutions. So this paper proposes best-first-search branch and bound (BFS-BB) to automatically generate the test data, and the branching and bounding operations have both been optimized.

Table 1: Some methods used in this paper

Name	Operation	Stage
Dynamic variable oredering (DVO)	Branching	State space search
Initial domain reduction (IDR)	Bounding	Initialization
Hill climbing (HC)	Branching	State space search

2.2 The search strategies

During the search process, variables are divided into three sets: past variables (short for PV, already instantiated), current variable (now being instantiated), and future variables (short for FV, not yet instantiated). BFS-BB includes two stages: initialization and state space search. Some methods in BFS-BB are described in Table 1. BFS-BB is described by pseudo-codes as follows.

Algorithm 1 BFS-BB

Input p:the path to be traversed

Output $result\{Variable \mapsto Value\}$:test data making p feasible

1: $result \leftarrow null$
2: $path\ infeasible \leftarrow true$
3: call **Algorithm 3. Initial domain reduction**
4: **if** $path\ infeasible = true$ **then**
5: **return** infeasible path
6: call **Algorithm 2. Dynamic variable ordering**
7: $V_1 \leftarrow$select (D_1)
8: initial state$\leftarrow(null,\ x_1,\ D_1,\ V_1,\ active)$
9: $S_{cur} \leftarrow$initial state
10: **while** $x_i \neq null$ **do**
11: call **Algorithm 4. Hill climbing**
12: **if** $S_{cur}=(Pre, x_i, D_i, V_i,\ inactive)$ **then**
13: backtrack
14: **else** $result \leftarrow result \bigcup \{x_i \mapsto V_i\}$
15: $FV \leftarrow FV\text{-}\{x_i\}$
16: $PV \leftarrow PV+\{x_i\}$
17: call **Algorithm 2. Dynamic variable ordering**
18: $V_i \leftarrow$select (D_i)
19: $S_{cur} \leftarrow (Pre, x_i, D_i, V_i,\ active)$
20: final state $\leftarrow S_{cur}$
21: **return** $result$

The first stage is to perform the initialization operations corresponding to lines 1 to 9. At first, IDR (Section 3.2.1) is used to partially reduce the input domains of all variables and find infeasible paths on occasion. All the variables in FV are permutated by DVO (Section 3.1) to form a queue and its head x_1 is determined the first variable to be instantiated. A value V_1 is selected from the domain of x_1 (D_1). With all these, the initial state is constructed as $(null,\ x_1,\ D_1,\ V_1,\ active)$, which is also the current state S_{cur}. Then the hill climbing (Section 3.2.2) process begins for x_1. For brevity, our following explanation refers to the hill climbing process for each x_i in FV.

In the state space search stage as shown by lines 10 to 21 , Hill climbing utilizes interval arithmetic to judge whether V_i for x_i leads to a conflict or not. In summary, the hill climbing process ends with two possibilities. One is that it finally finds the optima for x_i and reaches the peak of the hill, so the type of S_{cur} is changed into *extensive*, indicating that the local search for

x_i ends and DVO for the next variable will begin. The other is that it fails to find the optima for x_i and there is no more search space, so the type of S_{cur} is changed into *inactive*, indicating that the local search for x_i ends and backtracking is inevitable. BFS-BB ends when the hill climbing processes for all the variables succeed and there is no more variable to be permutated. All the *extensive* nodes make the solution path.

3 Optimized branching and bounding operations

3.1 optimized branching operation

This part focuses on the branching operation, which concerns the ordering of variables that determines the next variable to be instantiated. In our method, the next variable to be instantiated is selected to be the one with the minimal remaining domain size (the size of the domain after removing the values judged infeasible), because this can minimize the size of the overall search tree. The technique to break ties works when there are often variables with the same domain size. We use variables'ranks to break ties. In case of a tie, the variable with the higher rank is selected. This method gives substantially better performance than picking one of the tying variables at random. Rank is defined as follows.

Definition 1. Assuming that there are k branches along a path, the **rank** of a branch (n_{qa}, n_{qa+1}) (a∈[1, k]) marks its level in the sequence of the branches, denoted as rank(n_{qa}, n_{qa+1}).

The rank of the first branch is 1, the rank of the second one is 2, and the ranks of those following can be obtained analogously. The variables appearing on a branch enjoy the same rank as the branch. The rank of a variable on a branch where it does not appear is supposed to be *infinity*. As a variable may appear on more than one branch, it may have different ranks. The rule to break ties according to the ranks of variables is based on the heuristics from interval arithmetic that the earlier a variable appears on a path, the greater influence it has on the result of interval arithmetic along the path. Therefore, if the ordering by rank is taken between a variable that appears on the branch (n_{qa}, n_{qa+1}) and a variable that does not, then the former has a higher rank. That is because on the branch (n_{qa}, n_{qa+1}), the former has rank a while the latter has rank *infinity*. The comparison between a and *infinity* determines the ordering. The algorithm is described by pseudo-codes as follows.

Quicksort is utilized when permutating variables according to remaining domain size and returns Q_i as the result. If no variables have the same domain size, then DVO returns the head of Q_i (x_i). But if there are variables whose domain sizes are the same as that of the head of Q_i , then the ordering by rank is under way, which will terminate as soon as different ranks appear.

3.2 Optimized bounding operation

This part focuses on the bounding operation, which in fact is the optimization of interval arithmetic. For the purpose of improving efficiency, the optimized bounding operation is taken in both stages of BFS-BB.

Initial domain reduction

The optimized bounding operation taken in the initialization stage is used for initial domain reduction as well as infeasible path detection. Following is the introduction to the process that interval arithmetic functions. First we give the definition of branching condition.

Algorithm 2 Dynamic variable ordering

Input FV:the set of future variables
D_i:the domain of x_i($x_i \in FV$)
$(n_{qa}, n_{qa+1})(a \in [1, k])$:$k$ branches along the path
Output x_i:the selected variable to be next instantiated

```
1: Q_i ← quicksort(FV, |D_i|)
2: for i → 1:|Q_i| do
3:     if |D_i| ≠ |D_j| (j > i; x_i, x_j ∈ Q_i) then
4:         break
5:     else
6:         for (n_qa, n_qa+1)(a ∈ [1, k]) do
7:             if rank(n_qa, n_qa+1)(x_i) = rank(n_qa, n_qa+1)(x_j) then
8:                 a++
9:             elsepermutate x_i, x_j by rank(n_qa, n_qa+1)
10:                break
11: x_i ← head(Q_i)
12: return x_i
```

Definition 2. Let B be the set of Boolean values {true, false} and D^a be the domain of all variables before the a^{th} branch, if there are k branches along the path, the branching condition $Br(n_{qa}, n_{qa+1})$: $D^a \to B$ ($a \in [1, k]$)where n_{qa} is a branching node is calculated by formula (1).

$$Br(n_{qa}, n_{qa+1}) = \begin{cases} true, & D^a \bigcap \widetilde{D}^a \neq \emptyset; \\ false, & D^a \bigcap \widetilde{D}^a = \emptyset. \end{cases} \tag{1}$$

In formula (1), D^a satisfies all the a-1 branching conditions ahead and will be used as input for the calculation of the a^{th} branching condition, and $\widetilde{D}^a$which is a temporary domain is the result when calculating $Br(n_{qa}, n_{qa+1})$ with D^a and satisfies the a^{th} branching condition. $D^a \bigcap \widetilde{D}^a \neq \emptyset$ means that $D^a \bigcap \widetilde{D}^a$ satisfies all the a-1 branching conditions ahead and the a^{th} branching condition, ensuring that interval arithmetic can continue to calculate the remaining branching conditions, while $D^a \bigcap \widetilde{D}^a = \emptyset$ means that the path has been detected infeasible. In the initialization stage, following algorithm can be quite useful for infeasible path detection as well as initial domain reduction since it is quite clear that $D^1 \supseteq D^2 \ldots \supseteq D^k \supseteq D^{k+1}$.

Algorithm 3 Initial domain reduction

Input D^1:the input domain of all variables
Output D^{k+1}:the reduced domain of all variables

```
1: for i → 1:k do
2:     Br(n_qa, n_qa+1) ← false
3:     D̃^a ← calculate Br(n_qa, n_qa+1) with D^a
4:     if D^a ⋂ D̃^a ≠ ∅ then
5:         Br(n_qa, n_qa+1) ← true
6:         D^a+1 ← D^a ⋂ D̃^a
7:     else return
8: path infeasible ← false
9: return D^k+1
```

If the path is not infeasible, then D^{k+1} will be input as the domain of all variables for the state space search stage. If it is, because interval arithmetic analyzes the ranges of variables in

a conservative way, the path to be covered is no doubt infeasible, and it is unnecessary to carry out the following state space search.

Hill climbing

Hill climbing is used to judge whether a fixed value V_i for the current variable x_i makes path p feasible. In other words, a certain V_i that makes p feasible is the peak that we are trying to search for x_i. Initial value is very important for hill climbing. Our initial value selection strategy was introduced thoroughly in [16], which is quite effective. This part focuses on the process that interval arithmetic judges whether the value (V_i) assigned to a variable (x_i) leads to a conflict or not. Different from the initialization stage, a conflict detected in state space search stage implies that the current V_i for x_i will not lead the search in the right direction, but another V_i may not. So what we are trying to find in this conflict is the information useful for the next search step. To be exact, we seek to find some information to help reduce the domain of the current variable x_i (D_i). In accordance with the model of state space search, we give the following formula of the objective function $F(V_i)$, which is used to reduce the current domain of D_i.

$$F(V_i) = V_i - \sum_{a=1}^{k} (D^a \bigcap \widetilde{D}^a)(x_i) \tag{2}$$

The reduction of D_i is carried out according to the result of the objective function, a new V_i is selected from the reduced domain D_i, and interval arithmetic will restart to judge whether V_i causes a conflict. The above procedure is likened to climbing a hill. $F(V_i)$=0 implies that there is no conflict detected and V_i is the value judged to be appropriate for x_i. Otherwise D_i will again have to be reduced according to the return value of $F(V_i)$. In the procedure of hill climbing, the absolute value of $F(V_i)$ will approximate more closely to 0, which is the objective or the peak of the hill. The algorithm is shown by following pseudo-codes.

It can be seen that $F(V_i)$ provides both the upper and the lower bounds of D_i for its reduction, so the efficiency of the algorithm is improved greatly. After the conduction of Algorithm 4, the hill climbing for a variable(x_i) ends with two possibilities, which can be identified by *Type* of S_{cur}: *extensive* means that the peak is found and DVO for the next variable may continue, while *inactive* means that no search space is left for x_i, the peak is not found, and backtracking is inevitable.

3.3 Case study

In this part, we give a case study to explain in detail how BFS-BB works. An example with a program *test* and its corresponding CFG is shown in Figure 1. Adopting statement coverage, there is only one path to be traversed, namely, *Path*1: 0→1→2→3→4→5→6→7→8→9→10. We choose this example, because the constraints along the path are very strict for two variables. It is very obvious that $\{x_1 \mapsto 65, x_2 \mapsto 35\}$ is only one solution to the corresponding CSP.

For simplicity, the input domains of both variables are set [1, 100] with the size 100. Then IDR reduces them to $x1$:[31,99],$x2$:[1,69] as shown in Figure 2, which is used as input for the state space search stage. Then DVO works to determine the first variable to be instantiated as Table 2 shows, and $x1$ is picked out as shown in bold in the last column.

The initial value of $x1$ is selected from [31, 99] according to the method introduced in [16]. Assume that 80 is selected, then after three times of conflict and reduction to D_1, 65 is found for x_1 at last. $F(65)$=0 means that 65 is the peak of the hill that corresponding to x_1, which starts from 80. The hill climbing process is shown in Table 3. And since there is only one value 35 for $x2$, the search succeeds with $\{x_1 \mapsto 65, x_2 \mapsto 35\}$.

Algorithm 4 Hill climbing

Input D_i: the domain of x_i
V_i: the current value of x_i
Output S_{cur}:$(Pre, x_i, D_i, V_i,\ inactive)$when hill climbing succeeds
S_{cur}:$(Pre, x_i, D_i, V_i,\ extensive)$when hill climbing fails

1: **while** $|D_i|>1$ **do**
2: **for** $i \rightarrow 1{:}k$ **do**
3: $Br(n_{qa},\ n_{qa+1}) \leftarrow false$
4: $\widetilde{D}^a \leftarrow$calculate $Br(n_{qa}, n_{qa+1})$with D^a
5: **if** $D^a \bigcap \widetilde{D}^a \neq \emptyset$ **then**
6: $Br(n_{qa},\ n_{qa+1}) \leftarrow true$
7: $D^{a+1} \leftarrow D^a \bigcap \widetilde{D}^a$
8: **else** calculate $F(V_i)$
9: **break**
10: **if** $F(V_i) = 0$ **then**
11: $S_{cur} \leftarrow (Pre, x_i, D_i, V_i,\ extensive)$
12: **return** S_{cur}
13: **else**
14: **if** $F(V_i) < 0$ **then**
15: $D_i \leftarrow [V_i+1, V_i+F(V_i)]$
16: **else**$D_i \leftarrow [V_i\text{-}F(V_i), V_i\text{-}1]$
17: $V_i \leftarrow$select(D_i)
18: $S_{cur} \leftarrow (Pre, x_i, D_i, V_i,\ inactive)$
19: **return** S_{cur}

```
void test(int x1, int x2){
1:  if(x1>x2)
2:    if(x1+x2==100)
3:      if(x1>=20)
4:        if(x1-x2==30)
5:          printf("Solved! ");
}
```

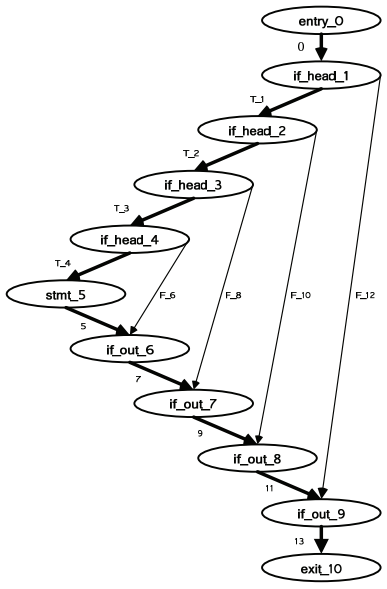


Figure 1: Program *test* and its corresponding CFG

Table 2: DVO process for $x1$ and $x2$

Ordering rule	Condition of each variable	Tie encountered?	Ordering result
Domain size	$\|D1\|$=69, $\|D2\|$=69	Yes	$\boldsymbol{x1}\rightarrow x2$
Rank1	Rank1($x1$)=1, Rank1($x2$)=1	Yes	
Rank2	Rank1($x1$)=2, Rank1($x2$)=2	Yes	
Rank2	Rank2($x1$)= 3, Rank2($x2$)=∞	No	

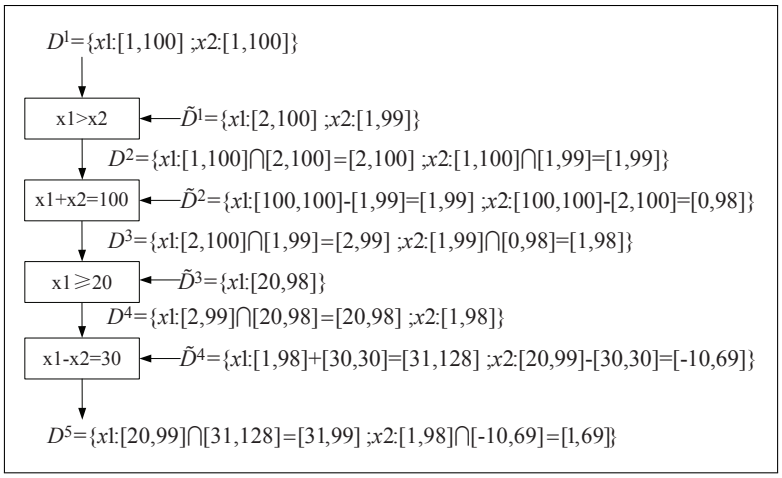


Figure 2: The IDR process

4 Experimental Analyses and Empirical Evaluations

To observe the effectiveness of BFS-BB, we carried out a large number of experiments in CTS. Within the CTS framework, the PUT is automatically analyzed, and its basic information is abstracted to generate its CFG. According to the specified coverage criteria, the paths to be traversed are generated and provided for BFS-BB as input. The experiments were performed in the environment of MS Windows 7 with 32-bits, Pentium 4 with 2.8 GHz and 2 GB memory. The algorithms were implemented in Java and run on the platform of eclipse.

4.1 Performance evaluation

The number of variables is an important factor that affects the performance of test data generation methods [17]. Hence, in this part, experiments were carried out to evaluate the effectiveness of the optimized branching and bounding operations for varying numbers of input variables.

Testing the optimized branching operation

This part presents the comparison between our branching algorithm DVO (A) which is also BFS-BB and the method which orders variables only by remaining domain sizes (B). The other operations (the bounding operations) in the two methods were both accomplished by IDR and HC. The comparison was accomplished by repeatedly running the two methods on generated test programs having input variables x_1, x_2,..., x_n where n varied from 2 to 50. Adopting statement coverage, in each test the program contained n *if* statements (equivalent to n branching conditions or n expressions along the path) and there was only one path to be traversed of fixed length, which was the one consisting of entirely true branches. In each test, the expression of

Table 3: The hill climbing process for $x1$

D_1	V_1	$F(V_i)$	$\lvert F(V_i)\rvert$	Peak reached?
[31, 99]	80	30	30	No
[50, 79]	60	-10	10	No
[61, 70]	65	0	0	Yes

the i^{th} $(1 \leq i \leq n)$ *if* statement was in the form of

$$[a_1, a_2, \ldots, a_n][x_1, x_2, \ldots, x_n]\prime rel - op \quad const[i] \tag{3}$$

In formula(3), a_1, a_2,..., a_i were randomly generated numbers either positive or negative, $rel\text{-}op \in \{>, \geq, <, \leq, =, \neq\}$, and $const[j]$ $(j \in [1, i])$ was an array of randomly generated constants within [0, 1000]. The randomly generated a_j and $const[i]$ should be selected to make the path feasible. This arrangement constructed the tightest linear relation between the variables. The programs for various values of n ranging from 2 to 50 were each tested 50 times, and the average time required to generate the data for each test was recorded. The results are presented in Figure 3.

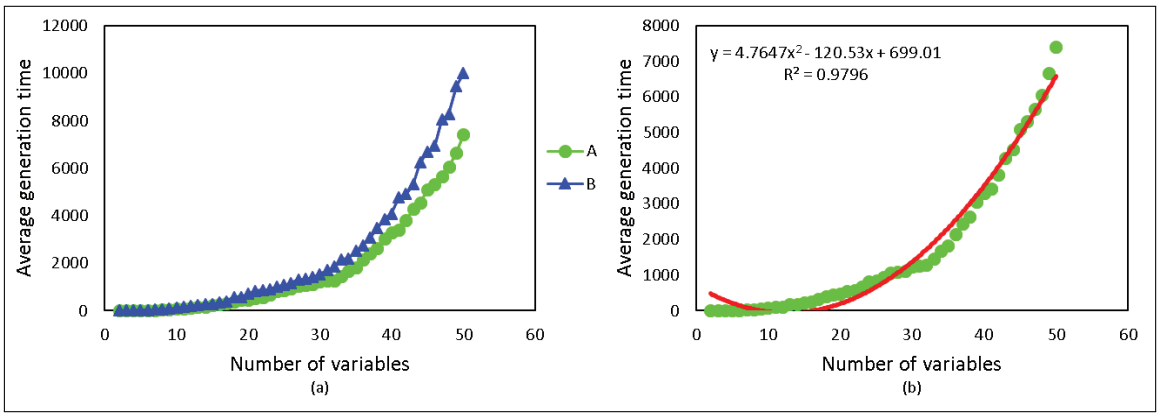


Figure 3: Test result of DVO

In Figure 3, (a) shows that A had a better performance than B, but it was not very obvious when the number of variables (expressions) was not very large, because there was no requirement for an optimized ordering algorithm, since remaining domain size was enough to determine the next variable to be instantiated. So the more variables, the better DVO works. For BFS-BB (A), it is clear that the relation between average generation time and the number of variables can be represented as a quadratic curve very well as shown in (b) and the quadratic correlation relationship is significant at 95% confidence level with p-value far less than 0.05. Besides, average generation time increases at a uniformly accelerative speed as the increase of the number of variables. The differentiation of average generation time indicates that its increase rate rises by y=9.5294x-120.53 as the number of variables increases. We can roughly draw the conclusion that generation time using DVO is very close for n ranging from 1 to 25, while it begins to increase when n is larger than 25. So DVO will be very useful for PUTs with more variables, especially the large-scale real-world programs.

Testing the optimized bounding operation

This part presents the comparison between our bounding algorithm HC (A) which is also BFS-BB and the method without HC (B). The other operations (DVO and IDR) in the two

methods were totally the same. The comparison was accomplished by repeatedly running the two methods on generated test programs having input variables x_1, x_2,..., x_n where n varied from 1 to 50. Adopting statement coverage, in each test the program contained 50 *if* statements (equivalent to 50 branching conditions or 50 expressions along the path) and there was only one path to be traversed of fixed length, which was the one consisting of entirely true branches. The expression of each *if* statement was in the form of

$$[a_1, a_2, \ldots, a_n][x_1, x_2, \ldots, x_n]\prime rel - op \quad const[c] \tag{4}$$

In formula(4), a_1, a_2,..., a_i were randomly generated numbers either positive or negative, $rel\text{-}op \in \{>, \geq, <, \leq, =, \neq\}$, and $const[c]$ ($c \in [1, 50]$) was an array of randomly generated constants within [0, 1000]. The randomly generated a_i ($1 \leq i \leq n$) and $const[c]$ should be selected to make the path feasible. This arrangement constructed the tightest linear relation between the variables. In addition, we ensured that there was at least one "=" in each program to test the equation solving capability of the methods. The programs for various values of n ranging from 1 to 50 were each tested 50 times, and the average time required to generate the data for each test was recorded. The comparison result is presented in Figure 4.

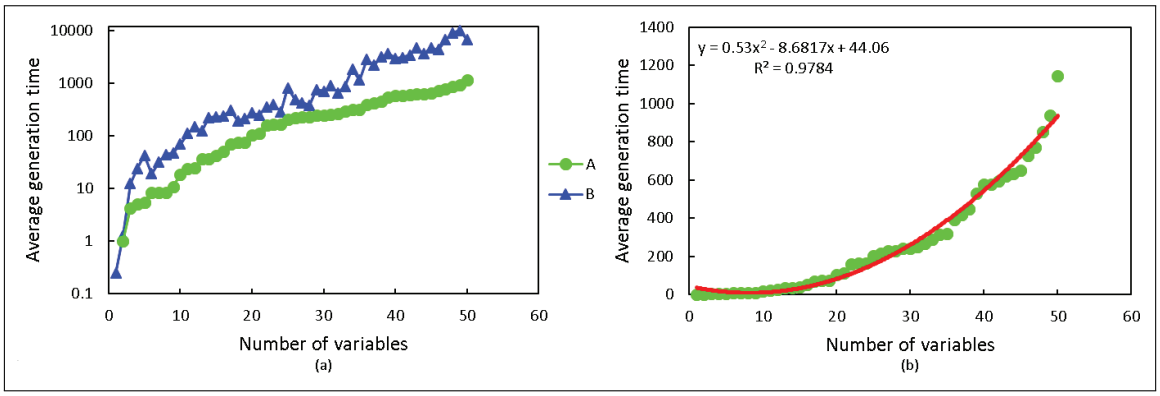


Figure 4: Test result of HC

We exponentiated the axis representing average generation time as shown in (a). It can be seen that the average generation time of A is far less than B. For BFS-BB (A), it is clear that the relation between average generation time and the number of variables can be represented as a quadratic curve very well as shown in (b) and the quadratic correlation relationship is significant at 95% confidence level with p-value far less than 0.05. Besides, average generation time increases at a uniformly accelerative speed as the increase of the number of variables. The differentiation of average generation time indicates that its increase rate rises by y=1.06x-8.6817 as the number of variables increases. We can roughly draw the conclusion that generation time using HC is very close for n ranging from 1 to 8, while it begins to increase when n is larger than 8.

4.2 Coverage evaluation

To evaluate the capability of BFS-BB to generate test data in terms of coverage, we used some real-world programs to compare BFS-BB with both static and dynamic methods adopted in test data generation.

Comparison with a static method

This part presents the results from an empirical comparison of BFS-BB with the static method [11] (denoted as "method 1" to avoid verbose description), which was implemented in CTS prior

Table 4: The details of comparison with method 1

Project	program	Function	AC by method 1	AC by BFS-BB
qlib	sin.c	radian	55%	**100%**
	floor.c	ceil	66%	**100%**
dell8i-2	asinl.c	acosl	55%	**100%**
	tanl.c	cotl	66%	**100%**

Table 5: Parameter setting for PSO

Parameter	Value
Population size	30
Max generations	100
Inertia weight w	Ranging from 0.2 to 1
Acceleration constants c_1 and c_2	$c_1 = c_2 = 2$
Maximum velocity V_{max}	Set according to the input space of the tested program

to BFS-BB. The test beds were from two engineering projects at http://www.moshier.net/. The comparison adopted statement coverage as the adequacy criterion. For each test bed, the experiments were carried out 100 times, and average coverage (AC) was used for comparison. The details of the comparison are shown in Table 4. From Table 4, it can be seen that BFS-BB reached higher coverage than method 1 for all the test beds as shown in bold. That is largely due to the optimization methods utilized in BFS-BB. The applicability in engineering remains one of our primary goals in the future.

Comparison with PSO

This part presents results from an empirical comparison of BFS-BB with PSO, which is mentioned in Section 1 as a popular MHS method with relatively fast convergence speed. Table 5 is a brief introduction to some parameters used in PSO. We used three real-world programs, which are well-known benchmark programs and have been widely adopted by other researchers [18]-[20]. Branch coverage was taken as the adequacy criterion. For each test bed, the experiments were carried out 100 times, and AC was used for comparison. Table 6 shows the details of the test beds and the comparison results. Obviously BFS-BB achieved 100% coverage as shown in bold on all the three benchmark programs, which are rather simple programs for BFS-BB, and it outperformed the algorithm in comparison. The better performance of BFS-BB is due to two factors. The first is that the initial values of variables are selected by heuristics on the path, so BFS-BB reaches a relatively high coverage for the first round of the search. The second is that the optimized bounding operation is conducted not only in the state space search stage but in the initialization stage as well, which reduces the domains of the variables to ensure a relatively small search space that follows.

Table 6: The details of comparison with PSO

Program	LOC	Branches	Variables	AC by PSO	AC by BFS-BB
triangleType	31	3	5	99.88%	**100%**
cal	53	18	5	96.85%	**100%**
calDay	72	11	3	97.35%	**100%**

5 Conclusions and Future Works

The increasing demand of testing large-scale real-world programs makes the automation of the testing process necessary. In this paper, path-wise test data generation (Q) which is a basic problem in software testing is defined as a constraint satisfaction problem (CSP), and the algorithm best-first-search branch and bound (BFS-BB) is presented to solve it, combining two techniques in artificial intelligence which are state space search and branch and bound (BB). The branching and bounding operations in BFS-BB are both optimized. For the branching operation, dynamic variable ordering (DVO) is proposed to permutate variables with a heuristic rule to break ties. The bounding operation is optimized in both stages of BFS-BB. Initial domain reduction (IDR) functions in the initialization stage to reduce the search space as well as detect infeasible paths. In the state space search stage, the process of determining a fixed value for a specified variable resembles climbing a hill, the peak of which is the value judged by interval arithmetic that does not cause a conflict. To facilitate the search procedure, the solution space is represented as state space. Empirical experiments show that the optimized branching operation is especially useful for large-scale programs, while the advantage of the optimized bounding operation hill climbing (HC) is very obvious. The results also show that BFS-BB outperforms some current static and dynamic methods in terms of coverage. Our future research will involve how to generate test data to reach high coverage. The effectiveness of the generation approach continues to be our primary work.

Acknowledgment

This work was supported by the National Grand Fundamental Research 863 Program of China (No. 2012AA011201), the National Natural Science Foundation of China (No. 61202080), the Major Program of the National Natural Science Foundation of China (No. 91318301), and the Open Funding of State Key Laboratory of Computer Architecture (No. CARCH201201).

Bibliography

[1] Michael R. Lyu; Sampath Rangarajan; Ada P. A. van Moorse. (2002); Optimal allocation of test resources for software reliability growth modeling in software development, *IEEE Transactions on Reliability*, ISSN 1841-9836, 51(2): 183-192.

[2] Tassey Gregory.(2002);The economic impacts of inadequate infrastructure for software testing, *National Institute of Standards and Technology, RTI Project 7007.011.*

[3] Weyuker Elaine J.(1999); Evaluation techniques for improving the quality of very large software systems in a cost-effective way, *Journal of Systems and Software*, ISSN 0164-1212, 47(2): 97-103.

[4] Shaukat Ali, Lionel C. Briand; Hadi Hemmati; Rajwinder K. Panesar-Walawege. (2010); A systematic review of the application and empirical investigation of search-based test case generation, *IEEE Transactions on Software Engineering*, ISSN 0098-5589, 36(6): 742-762.

[5] Mao Chengying; Yu Xinxin; Chen Jifu.(2012); Swarm Intelligence-Based Test Data Generation for Structural Testing, *Proceedings of 11th International Conference on Computer and Information Science (ICIS 12)*,623-628.

[6] Suzette Person; Guowei Yang; Neha Rungta; Sarfraz Khurshid. (2012); Directed incremental symbolic execution, *IACM SIGPLAN Notices*, ISSN 0362-1340, 46(6): 504-515.

[7] Moore Ramon Edgar; R. Baker Kearfott; Michael J. Cloud.(2009);*Introduction to interval analysis*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

[8] Richard A. DeMillo; A. Jefferson Offutt. (1991); Constraint-based automatic test data generation, *IEEE Transactions on Software Engineering*, ISSN 0098-5589, 17(9): 900-910.

[9] Arnaud Gotlieb; Bernard Botella; Michel Rueher.(1998);Automatic test data generation using constraint solving techniques, *Proceedings of the 1998 ACM SIGSOFT International Symposium on Software Testing and Analysis*, 23(2):53-62.

[10] Cristian Cadar; Daniel Dunbar; Dawson Engler.(2008);KLEE: Unassisted and automatic generation of high-coverage tests for complex systems programs, *Proceedings of USENIX Symposium on Operating Systems Design and Implementation (OSDI 2008)*, 209-224.

[11] Wang Yawen; Gong Yunzhan; Xiao Qing. (2013); A Method of Test Case Generation Based on Necessary Interval Set, *Journal of Computer-Aided Design & Computer Graphics*, ISSN 1003-9775, 25(4): 550-556.

[12] A.E. Eiben; Zs Ruttkay.(1997);*Constraint satisfaction problems*, pp. C5.7:1-8, New York, NY, USA: IOP Publishing Ltd and Oxford University Press.

[13] Ling-Ling Wang; Wen-Hsiang Tsai. (1988); Optimal assignment of task modules with precedence for distributed processing by graph matching and state-space search, *BIT Numerical Mathematics*, ISSN 0003-3835, 28(1): 54-68.

[14] Lianbo Gao; Shashi K. Mishra; Jianming Shi. (2012); An extension of branch-and-bound algorithm for solving sum-of-nonlinear-ratios problem, *Optimization Letters*, ISSN 1862-4472, 6(2): 221-230.

[15] Ying Xing; Junfei Huang; Yunzhan Gong; Yawen Wang; Xuzhou Zhang. (2014); An Intelligent Method Based on State Space Search for Automatic Test Case Generation, *Journal of Software*, ISSN 1796-217X, 9(2): 358-364.

[16] Ying Xing; Junfei Huang; Yunzhan Gong; Yawen Wang; Xuzhou Zhang. (2014); Path-wise Test Data Generation Based on Heuristic Look-ahead Methods, *Mathematical Problems in Engineering*, ISSN 1024-123X,volume 2014, Article ID 642630.

[17] Matthew J Gallagher; V. Lakshmi Narasimhan. (2012); Adtest: A test data generation suite for Ada software systems, *IEEE Transactions on Software Engineering*, ISSN 0098-5589, 23(8): 473-484.

[18] Mao Chengying; Yu Xinxin; Chen Jifu.(2012); Generating Test case for Structural Testing Based on Ant Colony Optimization, *Proceedings of the 12th International Conference on Quality Software (QSIC12)*, 98-101.

[19] Ammann Paul; Jeff Offutt.(2008); *Introduction to Software Testing*, Cambridge University Press, New York, NY, USA.

[20] E. Alba; F. Chicano. (2008); Observation in Using Parallel and Sequential Evolutionary Algorithms for Automatic Software Testing, *Computers & Operators Research*, ISSN 0305-0548, 35(10): 3161-3183.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(4):510-523, August, 2014.

Study on Directed Trust Graph Based Recommendation for E-commerce System

H.D. Zhong, S. Zhang, Y. Wang, Y. Shu

Haidong Zhong
Institute of Modern Logistics, Zhejiang Wanli University
No. 8 South Qianhu Rd., Ningbo, Zhejiang, 315100, PR, China
zhong_hai_dong@163.com

Shaozhong Zhang
Institute of Electronics and Information, Zhejiang Wanli University
No. 8 South Qianhu Rd., Ningbo, Zhejiang, 315100, PR, China
dlut_z88@163.com

Yanling Wang
Institute of Modern Logistics, Zhejiang Wanli University
No. 8 South Qianhu Rd., Ningbo, Zhejiang, 315100, PR, China
yanling_wang1@yahoo.com.cn

Yonggang Shu*
School of Tourism and City Management, Zhejiang Gongshang University
Xuezheng Street, Hangzhou, 310018, PR, China
*Corresponding author: zhd_1981@163.com

Abstract: Automated recommender systems have played a more and more important role in marketing and ever increasingly booming e-commerce systems. They provide useful predictions personalized recommendations according to customers characteristics and a variety of large and complex product offerings. In many of these recommendation technologies Collaborative Filtering (CF) has proven to be one of the most successful recommendation method, which has been widely used in many e-commerce systems. The success of CF recommendation depends mainly on locating similar neighbors to get recommendation items. However, many scholars have found that the process of finding similar neighbors often fail, due to some inherent weaknesses of CF based recommendation. In view of this, we propose a trust feedback recommendation algorithm based on directed trust graph (DTG), which is able to propagate trust relationship. In our approach, there is no need to compute similarity between users, but utilize the trust relation between them to conduct prediction calculation. Based on the analysis of human trust perception, we incorporate the process into our recommendation algorithm. Experimental evaluation on real life Epinions datasets shows that the effectiveness and practicability of our approach.
Keywords: Trust, Recommendation, Graph, E-commerce, Feedback.

1 Introduction

With the fast development of networking systems, an ever-increasing number of merchants are attracted to swarm into e-Commerce all over the world [1]. Accordingly, the Internet is changing from generally simple information exchange and extraction tool into the biggest virtualized market space presenting a great number of commercial services, ranging from electronic web-stores, on-line booking and service center to other social services [2]. With the popularization of the Internet, the amount of available information are exponentially growing and e-commerce systems structure becomes more complicated when it provides more and more choices for users [3, 4]. Under the circumstances, perform transaction tasks typically involved in e-Commerce, customers

have often to spend a large amount of time navigating among thousands of web pages to explore desired products and make their purchases. On the other hand, electronic business suppliers suffer from the problems of promoting their commodities to their potential customers in an effective way, considering their preferences, habits and other personalized characteristics. To deal with this challenge, researchers have advanced a recommendation approach which automatically analyze and mining e-commerce system visitors trading and browsed items data to filter web page information, classified newsgroup messages, and recommend valuable merchandise items [5].

Recommender systems, one of the most important computer-based intelligent approaches to find out the most appropriate services or goods from a large amount of products, are proved to be important tools that overcome the information overload by sifting through the large set of data and recommending information relevant to the user [6–10]. Typically, in e-commerce environment a recommender system analyzes trading data between consumer and sellers and items to find associations among them, and the items bought by similar users are presented as recommendations. Using this technology, some e-commerce systems, such as Amazon.com, Ebay.com and Netflix.com, are reported to have enhanced e-commerce sales by transforming e-commerce system browsers to buyers, increasing cross-selling and building customer loyalty [11].

As early as in the early 1990s intensive studies have been conducted in recommender systems, and many scholars deem them as knowledge discovery in database (KDD) systems or electronic agent systems [12–14]. Up to now, the existing recommendation means can be generally classified as content based, collaborative, knowledge-based, demographic, and utility based [15–17], among which collaborative filtering based personalized recommendation is proved to be one of the most successfully used technology [15]. In the newer, narrower sense, collaborative filtering builds correlations between pairs of items by making automatic predictions (filtering) about the interests of a user, and then figure out recommendations by finding items with high similarity to many users (collaborating). The acquiescent assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue T than to have the opinion on T of a person chosen randomly. For example, a collaborative filtering recommendation system for laptop preferences could make forecasts about which laptop a user should like given a partial list of that user's preferences (likes or dislikes) [10]. However there are two main inherent weaknesses in the collaborative filtering based recommendation systems [8]: (1) It is a challenge to find similar user, because the probability of two random users have rated any items in common is very small, and hence they are hardly comparable. (2) If ad hoc user profiles with the goal of being considered as similar to the target user a created, CF based recommendation can be easily attacked and recommendation precision can be greatly influenced.

In order to overcome these shortages, Massa P. and Avesani P. proposed a trust-aware recommendation approach [8]: Make use of trust propagation to search for trustable users based on the trust network instead of finding similar users as collaborative filtering does. The products bought or preferred by these users are then recommended to the active user over the trust network. Three years later, Ray S. and Mahanti A. put forward a new idea to improve prediction accuracy for trust aware recommender systems by removing all the trust statements that fall below a threshold correlation value to reconstruct the trust network [18]. They assume that a trust statement passed between two users should imply that similarity between both users will be relatively high, and generally utilize all the trust statements present in the data of similar users, calculated based on ratings, to make predictions. Experiment on Epinions datasets shows their method has better performance and effectiveness than that of the original approach for different levels of trust propagation and threshold correlation values [18]. However, every trust statement passed between users does not imply the correlation between them will also be high, because one user may pass trust statements to another user on the basis of perceived notion that

his (or her) predilections match with others, while similarity calculated based on ratings may show that they are different.

In our opinion, in trust based recommendation method users similarity calculating is not necessary. And we propose an approach where we construct a directed trust graph of users without considering the similarity between them. Based on the DTG we present a Trust Feed Back Recommendation Algorithm (TFBRA) to make recommendations for a user. It shows a substantial good performance for generating predictions through experimental evaluation on Epinions datasets [19].

The rest of the paper is organized as fellows. In section 2 we review some existing techniques, which are related to our work. In Section 3 the approach of trust based recommendation is explained in detail. Section 4 conducts verifying experiment, using real-life Epinions datasets, and the results of our evaluations is discussed. The last section draws conclusions and points out the related possible future work.

2 Related work

2.1 Trust

Since the birth of human beings and human social interactions trust came into being, and almost every aspect of a persons life is based on some form of trust. Undoubtedly, trust is positive and vital to humanity since it is part of love and friendship, and meaningful relationships depend upon it. Presently, trust is a research interest of many disciplines including management[18], marketing [21] and information systems [22]. However, scholars have difficult in reaching a consensus what exactly trust is, and they disagree even on the basic definitions. Presumptively, definitions of trust generally indicate a situation with characteristics of the following aspects [23–25]: One party (trustor) is willing to depend on the actions of another party (trustee); the situation is directed to the future. Moreover, the trustor (voluntarily or forcedly) abandons control over the actions performed by the trustee. As a consequence, the trustor is uncertain about the outcome of the other's actions; he can only develop and evaluate expectations. The uncertainty involves the risk of failure or harm to the trustor if the trustee will not behave as desired.

The rapid expansion of e-commerce conducts the research of trust in social science to new challenges. At the same time increasing the importance of trust and the urgency to know what customers shopping decision or preference [26], such as, interpersonal relationship between customers and sellers has been dis-intermediated by the method, and had to be improved upon. Additionally, e-commerce systems should be well designed or be made to imply the sellers are trustable, even if the regardless of seller's actual trustworthiness is not high. Many researchers have intensively studied the structure and formation mechanisms of trust from the aspects of both individual and organizational, and have identified five types antecedents to consumer trust, including institution-based, personality-based and calculation-based sources for trust building [27–30].

2.2 Recommendation System

Since the first appearance of the first paper on collaborative filtering in the mid-1990s, recommender systems have attracted many scholars' attention, and become an important research areas [31]. Over the last decade new approaches have been proposed to improve the efficiency and practicability of recommender system both in the industry and academia. At present recommender system related researches are still popular issues because they constitute problem-rich

research area and practical use requirement that help the potential buyers to solve information overload by providing personalized recommendation, and useful services according to their characteristics.

Over the past decades a lot of research work has been conducted on recommendation technologies (or algorithms), which used a wide range of statistical, artificial intelligence, information science and other techniques. These researches have observably improved the state-of-art in comparison to the previous recommender systems which used collaborative- and content-based heuristics. Up to now, Algorithmic means adopted in recommender systems can be divided into (1) content-based recommendation, collaborative, or hybrid, based recommendation and (2) heuristic-based or model-based recommendation based on the types of recommendation approaches used for the rating evaluation. Some of these methods are utilized in the industrial-strength recommender systems, e.g. the ones deployed at Amazon [32], MovieLens [33] and VERSIFI Technologies. However, despite there are many recommendation means, the present used recommender systems still need to be intensively improved, including better approaches for representing the recommended product items, more advanced and efficient recommendation algorithm or methods, utilization of multi-criteria ratings, to make recommendation methods more effective and applicable.

In recent years, scholars proposed a new recommendation method: trust based recommendation [8,34,35], and proved it more robust against shilling attacks and more capable and effective in generating recommendations for e-commerce system visitors, however they still need to calculate users similarity. Trust based recommendation systems are proved to make more accurate recommendation compared with traditional systems, because they utilize a new concept of trust propagation over a trust network. In [8], it has been experimentally shown how trust based recommender system outperform traditional recommendation methods on dataset from Epinions.com.

3 Trust based recommendation algorithm

In this section we start by introducing basic notations about trust cognitive process and concept of DTG. Based on the analysis of human trust perception, we incorporate the process into our recommendation algorithm. And then we present the logical architecture of trust based recommendation approach.

3.1 Trust cognitive analysis

From the viewpoint of recommendation in human society, it is easy to find that a person is usually more inclined to trust the recommendation information from his or her "acquaintance" than that of "strangers". In e-commerce recommendation systems all the people can be connected by trust relationship, which can be denoted as a DTG. As shown in Figure 1 (a), solid line represents the direct trust relationship between users (N_1 trustN_2, N_2 trustN_3). Through the trust relationship of N_2 andN_3, N_1 trust N_3 indirectly. If there is no trust relation between N_2 andN_3, it is impossible to create the trust relationship between N_1 and N_3 (As is shown in Figure (b)). Here we first define some basic trust related definitions in graph theory

Definition 1 *(Direct Trust)* Direct trust can be defined as an triple$< i, j, DT_{i,j} >$,stand for the directed edge from node i point to node$j$,and direct trust value of node i to node j is defined as$DT_{i,j}$, which is a discrete integer value in $[1, 5]$, and a greater value represents a deeper trust degree.

Definition 2 *(Indirect Trust)* Indirect trust can be defined as an triple$< i, j, IDT_{i,j} >$,Let $i$and$j$stand for nodes in the trust graph, $i$is attainable to$j$through limited hop$H(H > 1, H \in$

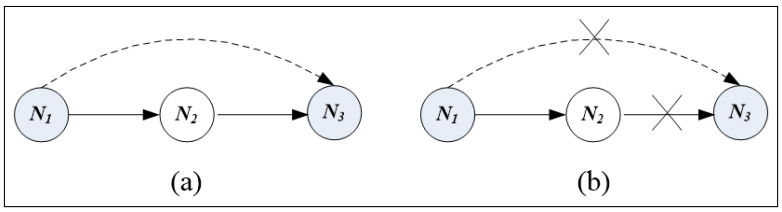


Figure 1: Trust process in human activity

N),and indirect trust value of node i to node j is defined as $IDT_{i,j}$, which is a discrete integer value in$[1, 5]$, and a greater value represents a deeper trust degree.

Theorem 1 *(Direct trust preference principle)* For any node i in the trust graph, its direct trust nodes are more credible than indirect trust nodes, where there are trust edges from i's direct trust nodes to its indirect trust nodes. For example, in Figure 1 (a) $DT_{N_1N_2}$ is more credible than IDT_{N_1N3} (or $DT_{N_1N_2} > IDT_{N_1N_3}$).

3.2 DTG based trust feedback

On the base of trust perception process presented above, we propose a DTG based trust feedback algrithm (The general principle is shown in Figure 2).

In figure 2 the edges represents the trust relationship between different nodes. While asking for recommendations of $\boldsymbol{U}_0$, the node just send a trust query information to its direct trust nodes ($\boldsymbol{U}_1$,$\boldsymbol{U}_3$,$\boldsymbol{U}_{15}$ and $\boldsymbol{U}_{16}$), and these nodes also send the trust query information to their trusted node similarly, until the end of trust inquiry control process and then all the queried node feedback their trust value of any products. Through the trust feedback process, comprehensive score of the items recommended by trusted node (include both direct trust nodes and indirect trust nodes) for $\boldsymbol{U}_0$.

3.3 DTG establishment

As narrated before, direct trust node and indirect nodes are used to represent the trust relationship based on graph theory. The step of establishing a DTG is instroduced here in detail below.

Definition 3 *(Direct Trust Node, DTN)* DTN of a certain node refers to any node that has direct trust relationship with it.

Definition 4 *(Indirect Trust Node, IDTN)* IDTN of a certain node refers to any node that has indirect trust relationship with it.

In this paper we adopt a five level value: 1,2,3,4,5 to repersent the trust(driect trust and indriect trust) value between one another. The higher trust value implies a further trust. In our framework the relationship between any node and its neighbor can be devided into three categories: direct trust relationship, indirect trust relationship and irrelevant nodes. For example, in table 1, U_0 has 4 DTNs. Figure 3 shows the process of establishing a DTG.

3.4 DTG based feedback trust recommendation fusion calculation

According to the transitivity attribute of trust, recommendated items for a specific node can be caculated on the base of feedback trust in DTG. If node N_i trust node N_j, N_i is more likely to accept the items that N_j gives a high rating value. And highly-rated-value items of DTNs of N_i account for a larger proportion in recommendated items than that of IDTNs of N_i.

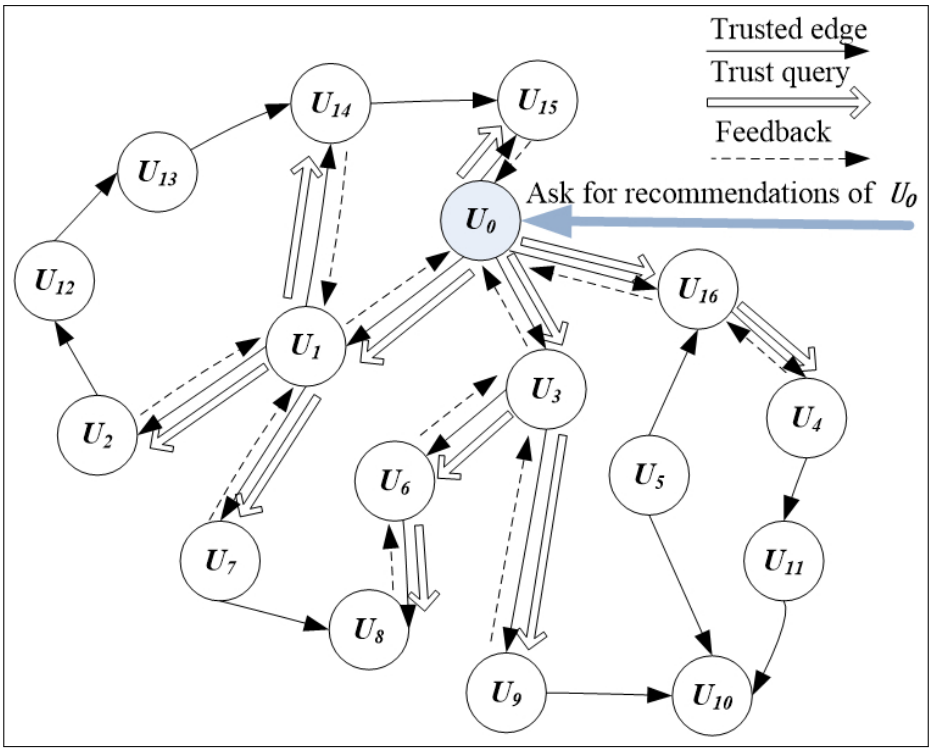


Figure 2: DTG based trust feedback principle

Table 1: Relationship between U_0 and its neighors

Peer	Neighbors	DTN
U_0	U_1	yes
U_0	U_3	yes
U_0	U_{15}	yes
U_0	U_{16}	yes

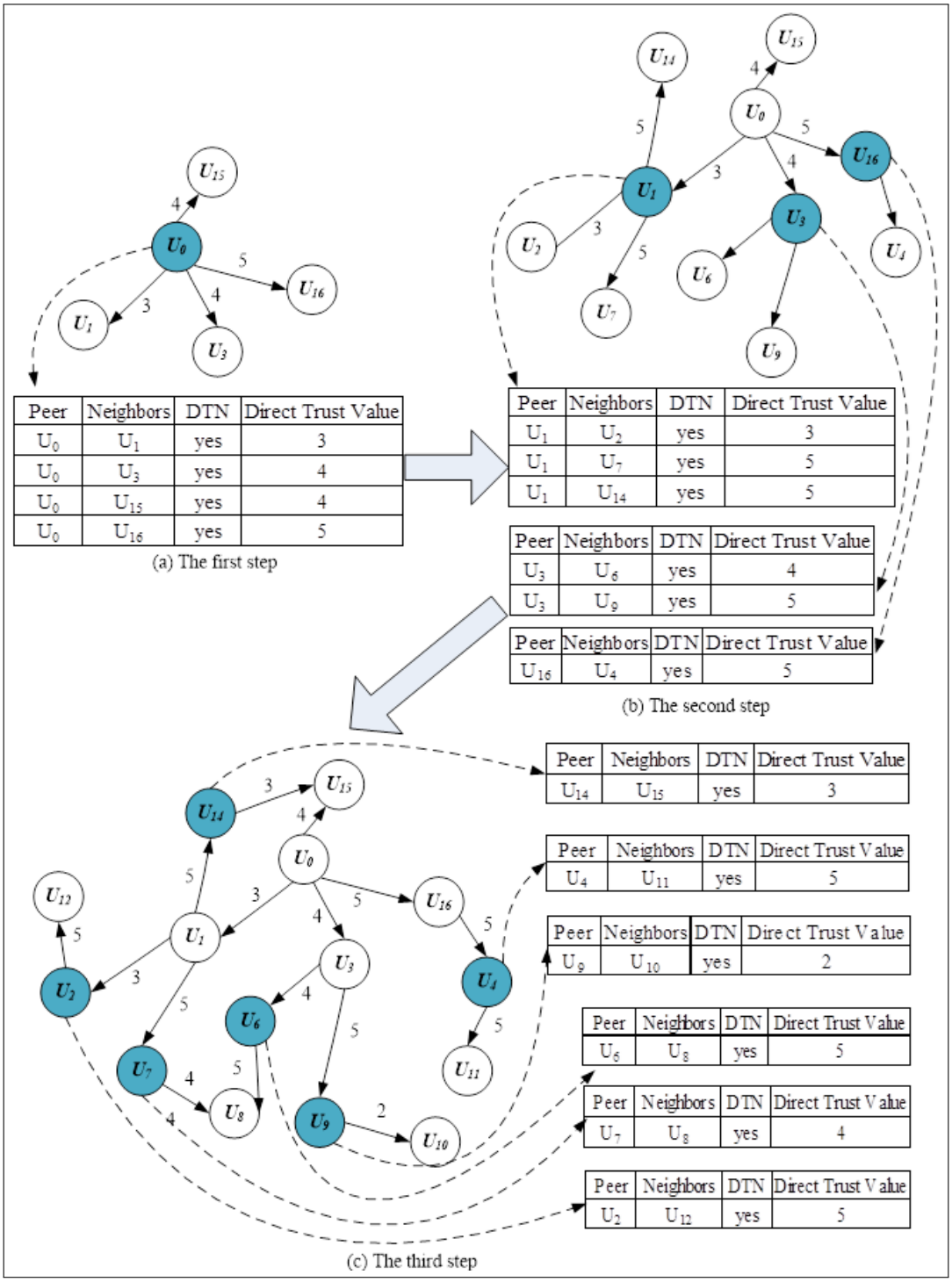


Figure 3: The process of establishing a DTG

Definition 5 The recommendated items $PRV(i)$of a certain node N_i can be abtained by comprehensive evaluation on the feedback trust value and high-rated-value items of its IDTNs and DTNs. Recommedated items from the trusted nodes(include DTNs and IDTNs) of N_i can be defined as the fellowing formula:

$$PRV(i) = \frac{\sum_{l=1}^{Level} W(l,\alpha) \sum_{v \in TNL_l} T_{sv} * R_{vi}}{\sum_{l=1}^{Level} W(l,\alpha) \sum_{v \in NTL_l} T_{sv}} \tag{1}$$

In the formula *Level* denotes the distance from node N_i,$W(l,\alpha)$ denotes weighted control factor function of the trust feedback value, which is related to the *Level* and attenuation factor α. In our opinion $W(l,\alpha)$ be defined as the fellowing formula:

$$W(l,\alpha) = \begin{cases} \alpha & , Level = 1 \\ \alpha \prod_{l=1}^{Level} \frac{1}{l} & , Level > 1 \end{cases} \tag{2}$$

In the formula α denotes the trust attenuation factor or initial trust factor.

3.5 Trust Feedback Recommendation Algorithm

As it has already been stated above, in e-commerce trading process people are more likely to accept the recommendations from direct trust persons. In order to simulate the trust based recommendation idea, we designed the trust feedback recommedation algrithm, the pseudo-code is as follows.

//TFBRA Pseudo-code

TrustFeedBackRecommedationAlgrithm

Input: *s, MaxLevel,* α /* s denotes the active user needs for recommendations, MaxLevel denotes maxium search level, α denotes attenuation factor*/

Output: PR /*Recommended items list with high predict rating value */

Begin

While $l <= MaxLevel$ ***do***

Search DTNs in the lth level in DTG

if(nodes in searched DTNs of s is not in the trust nodes list TNL_l)

put the nodes into a trust nodes list TNL_l

endif

search the DTNs in TNL_l

if$(l > MaxLevel)$

endwhile

else

l add-self

endwhile

$PRV(i) = \frac{\sum_{l=1}^{Level} W(l,\alpha) \sum_{v \in TNL_l} T_{sv} * R_{vi}}{\sum_{l=1}^{Level} W(l,\alpha) \sum_{v \in NTL_l} T_{sv}}$/* calcuate the predict rating value of each item of users in TNL, let $PRV(i)$ represent the predict rating value of item i*/

put the items with highest predicted rating value (PRV) into PR

End

4 Experimental evaluation and result discussion

4.1 Experiment Dataset

In order to examine the effectiveness of our algorithm, we perform our experiment on dataset from Epinions. The Epinions dataset was collected by Paolo Massa in a 5-week crawl (November/December 2003) from the Epinions.com web site. The dataset contains 49,290 users who rated a total of 139,738 different items at least once, writing, 664,824 reviews and 487,181 issued trust statements. Users and Items are represented by anonymized numeric identifiers [19].

The Epinions datasets contain two files: (1) ratings_data.txt.bz2 (2.5 Megabytes), which contains the ratings given by users to items. In this file every line has the format "user_id item_id rating_value". User_id is an integer in [1,49290], item_id is an integer in [1,139738] and rating_value is an integer in the range [1,5]; (2) trust_data.txt.bz2 (1.7 Megabytes), which contains the trust statements issued by users. In the file every line has the format "source_user_id target_user_id trust_statement_value". Source_user_id and target_user_id are integer in [1, 49290], and trust_statement_value is always 1 (since in the dataset there are only positive trust statements and not negative ones (distrust)).

4.2 Evaluation measures

To use the Epinions datasets in a more flexible way, we imported the two files (ratings_data.txt and trust_data.txt) into Microsoft SQL Server 2005 to create two tables (rating_data and trust_data). And we add a trust_value column into the trust_data table and set a random integer value in [1, 5] to represent the trust value between two users. Evaluation of the approach put forward in chapter 3 is conducted in our self-development recommendation prototype system, which is implemented on Microsoft Visual 2008 platform in Windows 7 Ultimate environment with a Intel Core i3-2310 2.1GHz (4 CPUs) Processor and 6 gigabyte memory. Main interface of the prototype system is shown in Figure 4.

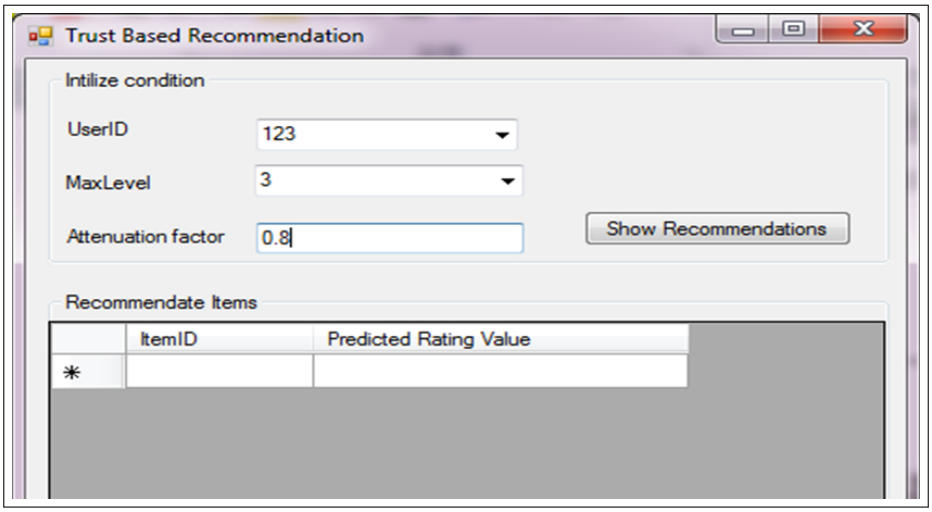


Figure 4: Main interface of the self-developed recommendation prototype system

4.3 Result and discussion

All the results are based on the self-developed trust based recommendation system. There are thousands of combination modes of the trust attenuation factor (α) with the maximum search

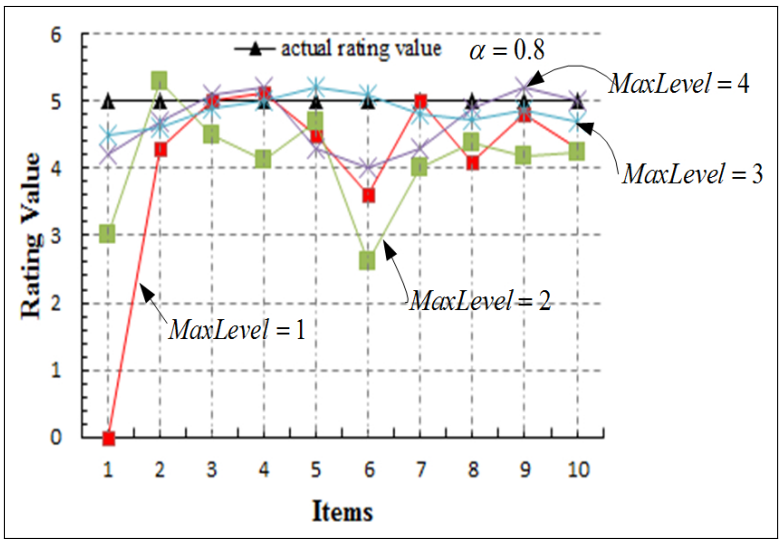


Figure 5: Ten recommended items for a randomly selected user are calculated according to TFBRA with αset to be 0.8

level ($MaxLevel$). We find it really hard to determine how α and $MaxLevel$ can be chosen. In order to prove the effectiveness of our recommend algorithm, two values (0.8 and 0.6) are chosen for parameterα, and ten recommended items for a randomly selected user are calculated according to TFBRA. Figure 5 shows the graphs for trust attenuation factor values with 0.8 and 0.6 respectively. Error between recommended value and actual rating value is shown in table 2. In the paper Mean Absolute Erro [36] (MAE) is used to evaluate the practical applicability of the trust feedback recommendation algorithm, a smaller MAE value means the higher accuracy. The calculation method of MAE is listed below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \qquad (3)$$

Where e_i is the difference between recommended rating value and actual rating value on the same item (product).

If a recommendation needed node have very few (or no) $DTNs$, MAE may be relatively great (As can be seen from Figure 5 and Figure 6 when$MaxLevel = 1$). With the increasing of$MaxLevel$,available feedback trust calculation nodes (include $DTNs$ and $IDTNs$) grows rapidly, which leads to more accurate recommendations. Although there are a great variety of combination modes of α with $MaxLevel$, it can be easily find from table 2 that MAE decrease obviously while $MaxLevel$ increase.

Table 2: Error between recommended rating value and actual rating value

MAE	MaxLevel=1	MaxLevel=2	MaxLevel=4	MaxLevel=5
$\alpha = 0.8$	0.954	0.957	0.41	0.221
$\alpha = 0.6$	0.757	0.49	0.4	0.331

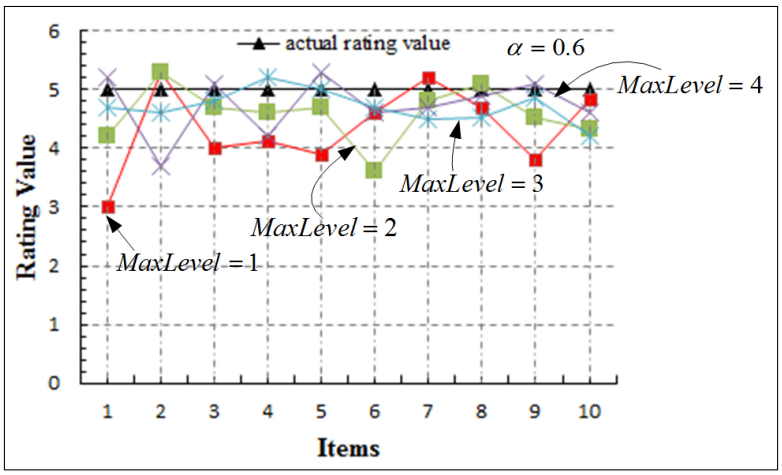


Figure 6: Ten recommended items for a randomly selected user are calculated according to TFBRA with αset to be 0.6

5 Conclusions

Recommendation is an important issue in information science; also it is a hot topic in e-commerce systems. When computing recommendations for users in e-commerce systems, most of the existing researches usually focus on find recommendations from similar users. However, many scholars have proved that there are many challenges in finding similar neighbors. The research tries to overcome the difficulties in traditional recommendation methods and put forward a new idea to get more to reasonable recommendations from trusted users. Based on the basic notations about trust cognitive process and DTG, which is constructed according to trust links between users, the paper proposes a trust feedback recommendation algorithm. We show through experiments on secondary processed Epinions datasets to better highlight the relative advantages of the different algorithms, and the experimental results indicate good effectiveness of our trust based recommendation approach.

One limitation of our method is that it has been test and verified on only one dataset i.e. Epinions datasets. In future we would like to prove the effectiveness of our trust feedback recommendation algorithm on trust datasets that exhibit characteristics different from Epinions datasets. Three important research questions that we would like to examine are: (1) Study the effect on recommendation accuracy when the way of setting trust value between two users varies; (2) Study the relation between weighted control factor function and the accuracy of our trust feedback recommendation algorithm; (3) Study the effectiveness of our approach using other datasets and compare the practicability of our method with other analogous means. We believe our approach can improve the coverage of recommender systems; however there is still much work need to be done.

Acknowledgements

This work is supported by the National Science Foundation of China (Grant No. 71373244), the Ningbo Natural Science Fundation project (Grant Nos. 2013A610068 and 2010A610119), the Zhejiang Province Department of Education project (Grant No. Y201329219), the Science Foundation of Zhejiang Province (Grant No. LY13G030033), the Modern Port Service Industry and Culture Research Center of the Key Research Base of Philosophy and Social Sciences of Zhejiang Province, the Science and Technology Innovation Team of Ningbo (Grant Nos. 2013B82009 and

2012B82003) and the Social Development Project of Ningbo (Grant No. 2012C50045).

Bibliography

[1] Papazoglou, M. P. (2001); Agent-oriented technology in support of e-business - Enabling the development of "intelligent" business agents for adaptive, reusable software, *Communications of the ACM*, ISSN 0001-0782, 44(4): 71-77.

[2] Levy, A. Y.; Weld, D. S. (2000); Intelligent Internet systems, *Artificial Intelligence*, ISSN 0004-3702, 118(1-2): 1-14.

[3] Peiguang, L.; Feng, Y.; Xiao, Y.; Qun, X. (2008); Personalized e-commerce recommendation based on ontology, *In Proc. of the International Conference on Internet Computing in Science and Engineering (ICICSE '08)*, 201-206.

[4] Siping H.; Meiqi F. (2008); Personalized recommendation based on ontology inference in E-commerce, *In Proc. of the International Conference on Management of e-Commerce and e-Government*, 192-195.

[5] Resnick, P.; Varian, H. R. (1997); Recommender systems, *Communications of the ACM*, ISSN 0001-0782, 40(3): 56-58.

[6] Huang, Z.; Chung, W. Y.; Chen, H. C. (2004); A graph model for e-commerce recommender systems, *Journal of the American Society for Information Science and Technology*, ISSN 1532-2882, 55(3): 259-274.

[7] Chen, L.; L'Abbate, M.; Thiel, U.; J. Neuhold, E. (2005); The Layer-Seeds Term Clustering Method: Enabling Proactive Situation-Aware Product Recommendations in E-Commerce Dialogues, *Information Systems Frontiers*, ISSN 1387-3326, 7(4/5): 405-419.

[8] Massa, P.; Avesani, P. (2007); Trust-aware Recommender Systems, *In Proc. of the Proceeding of ACM Recommender Systems Conference*, 17-24.

[9] Nan, R.; Qin, L. (2009); Research on the trust model based on the groups' internal recommendation in e-commerce environment, *Journal of Software Engineering and Applications*, ISSN 1945-3116, 2(4): 283-287.

[10] Konstan, J. A.; Riedl, J. (2012); Recommender systems: from algorithms to user experience, *User Modeling and User-Adapted Interaction*, ISSN 0924-1868, 22(1-2): 101–123.

[11] Zhang, W. (2007); A novel trust model based on recommendation for e-commerce, *In Proc. of the International Conference On Service Systems and Service Management*, 1-4.

[12] Ben Schafer, J.; Konstan, J.; Riedl, J. (1999); Recommender Systems in E-Commerce, *In Proc. of the ACM Conference on Electronic Commerce*, 158-166.

[13] Dzitac, I.; Barbat B.E. (2009); Artificial Intelligence + Distributed Systems = Agents, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 4(1):17-26.

[14] Zmaranda, D.; Silaghi, H.; Gabor, G.; Vancea, C. (2013); Issues on Applying Knowledge-Based Techniques in Real-Time Control Systems, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 4(1), 8(1): 166-175.

[15] Zuo, M.; Wang, K.; Li, J. (2005); The Application of Collaborative Filtering for Trust Management in P2P Communities, *Lecture Notes in Computer Science*, ISSN 0302-9743, 3758: 383- 394.

[16] Huang, Z.; Zeng, D. D. (2011); Why does collaborative filtering work?Transaction-based Recommendation Model Validation and Selection by Ananlyzing Bipartite Random Graphs, *Journal on Computing*, ISSN 2010-2283, 23(1): 138-152.

[17] Walter, F. E.; Battiston, S.; Schweitzer, F. (2008); A model of a trust-based recommendation system on a social network, *Autonomous Agents and Multi-Agent Systems*, ISSN 1387-2532, 16(1): 57-74.

[18] Ray, S.; Mahanti, A. (2010); Improving Prediction Accuracy in Trust-Aware Recommender Systems, *In Proc. of the 43rd Hawaii International Conference on System Sciences (HICSS)*, 1-9.

[19] http://www.trustlet.org/wiki/Downloaded_Epinions_dataset.

[20] Mayer, R. C.; Davis, J. H.; Schoorman, F. D. (1995); An Intergrative Model of Ogranization Turst, *Academy of Management Review*, ISSN 0363-7425, 20(3): 709-734.

[21] Morgan, R. M.; Hunt, S. D. (1994); The Commitment-Trust Theory of Relationship Marketing, *Journal of Marketing*, ISSN 0022-2429, 58(3): 20-38.

[22] Wang, Y. D.; Emurian, H. H. (2005); An overview of online trust: Concepts, elements, and implications, *Computers in Human Behavior*, ISSN 0747-5632, 21(1): 105-125.

[23] Schoorman, F. D.; Mayer, R. C.; Davis, J. H. (2007); An integrative model of organizational trust: past, present, and future, *Academy of Management Review*, ISSN 0363-7425, 32(2): 344-354.

[24] McKnight, D. H.; Chervany, N. L. (1996); The Meanings of Trust: University of Minnesota.

[25] Bamberger, W. (2010); Interpersonal Trust-Attempt of a Definition: Technische Universität München.

[26] McKnight, D. H.; Chervany, N. L. (2001); Conceptualizing trust: a typology and e-commerce customer relationships model, *In Proc. of the The 34th Annual Hawaii International Conference onSystem Sciences*, 36-44.

[27] McKnight, D. H.; Cummings, L. L.; Chervany, N. L. (1998); Initial trust formation in new organizational relationships, *The Academy of Management Review*, ISSN 0363-7425, 23(3): 473-490.

[28] Li, X.; Hess, T. J.; Valacich, J. S. (2008); Why do we trust new technology? A study of initial trust formation with organizational information systems, *The Journal of Strategic Information Systems*, ISSN 0963-8687, 17(1): 39-71.

[29] Gefen, D.; Karahanna, E.; Straub, D. W. (2003); Trust and TAM in online shopping: An integrated model, *MIS Quarterly*, ISSN 0276-7783, 27(1): 51-90.

[30] Kim, D. J.; Ferrin, D. L.; Rao, H. R. (2008); A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents, *Decision Support Systems*, ISSN 0167-9236, 44(2): 544-564.

[31] Resnick, P.; Iakovou, N.; Sushak, M.; Bergstrom, P.; Riedl, J. (1994); GroupLens: An open architecture for collaborative filtering of net news, *In Proc. of the Computer Supported Cooperative Work Conference*, 175-186.

[32] Linden, G.; Smith, B.; York, J. (2003); Amazon.com recommendations: item-to-item collaborative filtering, *IEEE Internet Computing*, ISSN 1089-7801, 7(1): 76-80.

[33] Miller, B. N. A.; Istvan, Lam, S. K.; Konstan, J. A.; Riedl, J. (2003); MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System, *In Proc. of the International Conference on Intelligent User Interfaces*, 263-266.

[34] Massa, P.; Avesani, P. (2004); *Trust-aware collaborative filtering for recommender systems*, In Meersman, R.; Tari Z.; VanderAalst W.; Bussler C.; Gal A.; Cahill V.; Vinoski S; Vogels W.; Gatarci T; Sycara K. (Eds.), *On the Move to Meaningful Internet Systems*, Springer-Verlag Berlin.

[35] Yuan, W., Shu, L.; Chao, H. C.; Guan, D.; Lee, Y. K.; Lee, S. (2010); ITARS: trust-aware recommender system using implicit trust networks, *IET Communications*, ISSN 1751-8628, 4(14): 1709-1721.

[36] Herlocker, J. L.; Konstan, J. A.; Terveen, K.; Riedl, J. T. (2004); Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems*, ISSN 1046-8188, 22(1): 5-53.

Author index