Advances in Intelligent Systems and Computing 298

Jeng-Shyang Pan · Vaclav Snasel
Emilio S. Corchado · Ajith Abraham
Shyue-Liang Wang *Editors*

Intelligent Data Analysis and Its Applications, Volume II

Proceeding of the First Euro-China Conference on Intelligent Data Analysis and Applications, June 13–15, 2014, Shenzhen, China

Springer

Advances in Intelligent Systems and Computing

Volume 298

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/11156

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

Jeng-Shyang Pan · Vaclav Snasel
Emilio S. Corchado · Ajith Abraham
Shyue-Liang Wang

Intelligent Data analysis and Its Applications, Volume II

Proceeding of the First Euro-China Conference on Intelligent Data Analysis and Applications, June 13–15, 2014, Shenzhen, China

Springer

Editors
Jeng-Shyang Pan
University of Applied Sciences
Kaohsiung
Taiwan

Vaclav Snasel
Department of Computer Science
Faculty of Elec. Eng. & Comp. Sci.
VSB-Technical University of Ostrava
Ostrava
Czech Republic

Emilio S. Corchado
Departamento de Informáca y Automática
Facultad de Biología
University of Salamanca
Salamanca
Spain

Ajith Abraham
Machine Intelligence Research Labs (MIR Labs)
Scientific Network for Innovation and Research Excellence
Washington
USA

Shyue-Liang Wang
Department of Information Management
National University of Kaohsiung
Kaohsiung
Taiwan

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-07772-7 ISBN 978-3-319-07773-4 (eBook)
DOI 10.1007/978-3-319-07773-4
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940737

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume composes the proceedings of the First Euro-China Conference on Intelligent Data Analysis and Applications (ECC 2014), which was hosted by Shenzhen Graduate School of Harbin Institute of Technology and was held in Shenzhen City on June13–15, 2014. ECC 2014 was technically co-sponsored by Shenzhen Municipal People's Government, IEEE Signal Processing Society, Machine Intelligence Research Labs, VSB-Technical University of Ostrava (Czech Republic), National Kaohsiung University of Applied Sciences (Taiwan), and Secure E-commerce Transactions (Shenzhen) Engineering Laboratory of Shenzhen Institute of Standards and Technology. It aimed to bring together researchers, engineers, and policymakers to discuss the related techniques, to exchange research ideas, and to make friends.

113 papers were accepted for the final technical program. Four plenary talks were kindly offered by: Ljiljana Trajkovic (IEEE SMC president), C.L. Philip Chen (IEEE Fellow, University of Macau), Jhing-Fa Wang (Tajen University, Taiwan), and Ioannis Pitas (University of Thessaloniki, Greece).

We would like to thank the authors for their tremendous contributions. We would also express our sincere appreciation to the reviewers, Program Committee members and the Local Committee members for making this conference successful. Finally, we would like to express special thanks for the financial support from Shenzhen Municipal People's Government and Shenzhen Graduate School of Harbin Institute of Technology in making ECC 2014 possible.

June 2014

Jeng-Shyang Pan
Vaclav Snasel
Emilio S. Corchado
Ajith Abraham
Shyue-Liang Wang

Conference Organization

Honorary Chairs

Han-Chieh Chao	National Ilan University, Taiwan

Advisory Committee Chairs

Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Bin-Yih Liao	National Kaohsiung University of Applied Sciences, Taiwan

Conference Chairs

Jeng-Shyang Pan	Harbin Institute of Technology Shenzhen Graduate School, China
Vaclav Snasel	VSB-Technical University of Ostrava, Czech Republic

Program Committee Chairs

Emilio S. Corchado	University of Salamanca, Spain
Ajith Abraham	Machine Intelligence Research Labs, USA
Shyue-Liang Wang	National University of Kaohsiung, Taiwan

Invited Session Chairs

Wu-Chih Hu	National Penghu University of Science and Technology, Taiwan
Kuo-Kun Tseng	Harbin Institute of Technology Shenzhen Graduate School, China

Electronic Media Chairs

Jiun-Huei Ho	Cheng Shiu University, Taiwan
Tsu-Yang Wu	Harbin Institute of Technology Shenzhen Graduate School, China

Local Organizing Chairs

Yanfeng Zhang	Harbin Institute of Technology Shenzhen Graduate School, China
Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China
Chien-Ming Chen	Harbin Institute of Technology Shenzhen Graduate School, China

Publication Chairs

Shu-Chuan Chu	Flinders University, Australia

Finance Chairs

Linlin Tang	Harbin Institute of Technology Shenzhen Graduate School, China

International Program Committee

Abdel hamid Bouchachia	University of Klagenfurt, Austria
Abd. Samad Hasan Basari	Universiti Teknikal Malaysia Melaka, Malaysia
Abraham Duarte	Universidad Rey Juan Carlos, Spain
Akira Asano	Kansai University, Japan
Alberto Alvarez	European Centre for Soft Computing, Spain
Alberto Cano	University of Cordoba, Spain
Alberto Fernandez	Universidad de Jaen, Spain
Alberto Bugarin	University of Santiago de Compostela, Spain
Alex James	Indian Institute of Information Technology and Management – Kerala, India
Alexandru Floares	Cancer Institute Cluj-Napoca, Romania
Alma Gomez	University of Vigo, Spain
Amelia Zafra Gomez	University of Cordoba, Spain
Amparo Fuster-Sabater	Institute of Applied Physics (C.S.I.C.), Spain
Ana Lorena	Federal University of ABC, Brazil
Anazida Zainal	Universiti Teknologi Malaysia, Malaysia
Andre Carvalho	University of Sao Paulo, Brazil
Andreas Koenig	Technische Universitat Kaiserslautern, Germany

Anna Bartkowiak	University of Wroclaw, Poland
Anna Fanelli	Universita di Bari, Italy
Antonio Peregrin	University of Huelva, Spain
Antonio J. Tallon-Ballesteros	University of Seville, Spain
Anusuriya Devaraju	Forschungszentrum Julich GmbH, Germany
Aranzazu Jurio	Universidad Publica de Navarra, Spain
Ashish Umre	University of Sussex, United Kingdom
Ashraf Saad	Armstrong Atlantic State University, United States
Ayeley Tchangani	University Toulouse III , France
Aymeric Histace	Universite Cergy-Pontoise, France
Azah Kamilah Muda	Universiti Teknikal Malaysia Melaka, Malaysia
Bartosz Krawczyk	Politechnika Wrocławska, Poland
Beatriz Pontes	University of Seville, Spain
Brijesh Verma	Central Queensland University, Australia
Carlos Barranco	Pablo de Olavide University, Spain
Carlos Cano	University of Granada, Spain
Carlos Fernandes	GeNeura Team, Spain
Carlos Garcia-Martinez	University of Cordoba, Spain
Carlos Lopezmolina	Universidad Publica de Navarra, Spain
Carlos Morell	Universidad Central Marta Abreu de Las Villas, Cuba
Cesar Hervas-Martinez	University of Cordoba, Spain
Chang-Shing Lee	National University of Tainan, Taiwan
Chao-Chun Chen	Southern Taiwan University, Taiwan
Chia-Feng Juang	National Chung-Hsing University, Taiwan
Chien-Ming Chen	Harbin Institute of Technology Shenzhen Graduate School, China
Chin-Chen Chang	Feng Chia University, Taiwan
Chris Cornelis	Ghent University, Belgium
Chuan-Kang Ting	National Chung Cheng University, Taiwan
Chu-Hsing Lin	Tunghai University, Taiwan
Chun-Wei Lin	Harbin Institute of Technology Shenzhen Graduate School, China
Coral del Val	University of Granada, Spain
Crina Grosan	Norwegian University of Science and Technology, Norway
Cristina Rubio-Escudero	University of Sevilla, Spain
Cristobal Romero	University of Cordoba, Spain
Cristobal J. Carmona	University of Jaen, Spain
Dalia Kriksciuniene	Vilnius University, Lithuania
David Becerra-Alonso	ETEA-INSA, Spain
Detlef Seese	Karlsruhe Institut of Technology (KIT), Germany
Edurne Barrenechea	Universidad Publica de Navarra, Spain
Eiji Uchino	Yamaguchi University, Japan

Eliska Ochodkova	VŠB Technical University of Ostrava, Czech Republic
Elizabeth Goldbarg	Federal University of Rio Grande do Norte, Brazil
Emaliana Kasmuri	Universiti Teknikal Malaysia Melaka, Malaysia
Enrique Herrera-Viedma	University of Granada, Spain
Enrique Yeguas	University of Cordoba, Spain
Eulalia Szmidt	Systems Research Institute Polish Academy of Sciences, Poland
Eva Gibaja	University of Cordoba, Spain
Federico Divina	Pablo de Olavide University, Spain
Fernando Bobillo	University of Zaragoza, Spain
Fernando Delaprieta	University of Salamanca, Spain
Fernando Gomide	University of Campinas, Brazil
Fernando Jimenez	University of Murcia, Spain
Francesc J. Ferri	Universitat de Valencia, Spain
Francesco Marcelloni	University of Pisa, Italy
Francisco Fernandez Navarro	University of Cordoba, Spain
Francisco Herrera	University of Granada, Spain
Francisco Martinez-Alvarez	Pablo de Olavide University, Spain
Francisco Martinez-Estudillo	University Loyola Andalucia, Spain
Frank Klawonn	University of Applied Sciences Baunschweig, Germany
Gabriel Luque	University of Malaga, Spain
Gede Pramudya	Universiti Teknikal Malaysia Melaka, Malaysia
Giacomo Fiumara	University of Messina, Italy
Giovanna Castellano	Universita di Bari, Italy
Giovanni Acampora	University of Salerno, Italy
Girijesh Prasad	University of Ulster, United Kingdom
Gladys Castillo	University of Aveiro, Portugal
Gloria Bordogna	CNR IDPA, Italy
Gregg Vesonder	AT&T Labs Research, United States
Huiyu Zhou	Queen's University Belfast, United Kingdom
Ilkka Havukkala	Intellectual Property Office of New Zealand, New Zealand
Imre Lendak	University of Novi Sad, Serbia
Intan Ermahani A. Jalil	Universiti Teknikal Malaysia Melaka, Malaysia
Isabel Nunes	UNL/FCT, Portugal
Isabel S. Jesus	Instituto Superior de Engenharia do Porto, Portugal
Ivan Garcia-Magarino	Universidad a Distancia de Madrid, Spain
Jae Oh	Syracuse University, United States
Jan Martinovic	VŠB Technical University of Ostrava, Czech Republic
Jan Plato	VŠB Technical University of Ostrava, Czech Republic

Javier Perez	University of Salamanca, Spain
Javier Sedano	Technological Institute of Castilla y Leon, Spain
Jesus Alcala-Fdez	University of Granada, Spain
Jesus Serrano-Guerrero	University of Castilla-La Mancha, Spain
Jitender S. Deogun	University of Nebraska, United States
Joaquin Lopez Fernandez	University of Vigo, Spain
Jorge Nunez Mc Leod	Institute of C.E.D.I.A.C, Argentina
Jose Luis Perez de la Cruz	University of Malaga, Spain
Jose M. Merigo	University of Barcelona, Spain
Jose-Maria Luna	University of Cordoba, Spain
Jose Pena	Universidad Politecnica de Madrid, Spain
Jose Raul Romero	University of Cordoba, Spain
Jose Tenreiro Machado	Instituto Superior de Engenharia do Porto, Portugal
Jose Valente De Oliveira	Universidade do Algarve, Portugal
Jose Villar	Oviedo University, Spain
Juan Botia	Universidad de Murcia, Spain
Juan Gomez-Romero	Universidad Carlos III de Madrid, Spain
Juan Vidal	Universidade de Santiago de Compostela, Spain
Juan J. Flores	Universidad Michoacana de San Nicolas de Hidalgo, Mexico
Juan-Luis Olmo	University of Cordoba, Spain
Julio Cesar Nievola	Pontificia Universidade Catolica do Parana, Brazil
Jun Zhang	Waseda University, Japan
Jyh-Horng Chou	National Kaohsiung First Univ. of Science and Technology, Taiwan
Kang Tai	Nanyang Technological University, Singapore
Kaori Yoshida	Kyushu Institute of Technology, Japan
Kazumi Nakamatsu	University of Hyogo, Japan
Kelvin Lau	University of York, United Kingdom
Kubilay Ecerkale	Turkish Air Force Academy, Turkey
Kumudha Raimond	Karunya University, India
Kun Ma	University of Jinan, China
Leandro Coelho	Pontificia Universidade Catolica do Parana, Brazil
Lee Chang-Yong	Kongju National University, Korea
Leida Li	University of Mining and Technology, China
Leon Wang	National University of Kaohsiung, Taiwan
Liang Zhao	University of Sao Paulo, Brazil
Liliana Ironi	IMATI-CNR, Italy
Luciano Stefanini	University of Urbino “Carlo Bo”, Italy
Ludwig Simone	North Dakota State University, United States
Luigi Troiano	University of Sannio, Italy
Luka Eciolaza	European Centre for Soft Computing, Spain
Macarena Espinilla Estevez	Universidad de Jaen, Spain
Manuel Grana	University of Basque Country, Spain

Manuel Lama	Universidade de Santiago de Compostela, Spain
Manuel Mucientes	University of Santiago de Compostela, Spain
Marco Cococcioni	University of Pisa, Italy
Maria Nicoletti	Federal University of Sao Carlos, Brazil
Maria Torsello	Universita di Bari, Italy
Maria Jose Del Jesus	Universidad de Jaen, Spain
Mariantonietta Noemi La Polla	IIT-CNR, Italy
Maria Teresa Lamata	University of Granada, Spain
Mario Giovanni C.A. Cimino	University of Pisa, Italy
Mario Koeppen	Kyushu Institute of Technology, Japan
Martine De Cock	Ghent University, Belgium
Michael Blumenstein	Griffith University, Australia
Michal Kratky	VŠB Technical University of Ostrava, Czech Republic
Michal Wozniak	Wroclaw University of Technology, Poland
Michela Antonelli	University of Pisa, Italy
Mikel Galar	Universidad Publica de Navarra, Spain
Milos Kudelka	VŠB Technical University of Ostrava, Czech Republic
Min Wu	Oracle, United States
Noor Azilah Muda	Universiti Teknikal Malaysia Melaka, Malaysia
Norberto Diaz-Diaz	Pablo de Olavide University, Spain
Norton Gonzalez	University of Fortaleza, Brazil
Nurulakmar Emran	Universiti Teknikal Malaysia Melaka, Malaysia
Olgierd Unold	Wroclaw University of Technology, Poland
Oscar Castillo	Tijuana Institute of Technology, Mexico
Ovidio Salvetti	ISTI-CNR, Italy
Ozgur Koray Sahingoz	Turkish Air Force Academy, Turkey
Pablo Villacorta	University of Granada, Spain
Patrick Siarry	Universit de Paris, France
Paulo Carrasco	Universidade do Algarve, Portugal
Paulo Moura Oliveira	University of Tras-os-Montes and Alto Douro, Portugal
Pedro Gonzalez	University of Jaen, Spain
Philip Samuel	Cochin University of Science and Technology, India
Pierre-Francois Marteau	Universite de Bretagne Sud, France
Pietro Ducange	University of Pisa, Italy
Punam Bedi	University of Delhi, India
Qieshi Zhang	Waseda University, Japan
Qinghan Xiao	Defence R&D Canada, Canada
Radu-Codrut David	Politehnica University of Timisoara, Romania
Rafael Bello	Universidad Central de Las Villas, Cuba
Ramin Halavati	Sharif University of Technology, Iran

Ramiro Barbosa	Instituto Superior de Engenharia do Porto, Portugal
Ramon Sagarna	University of Birmingham, United Kingdom
Richard Jensen	Aberystwyth University, United Kingdom
Robert Berwick	Massachusetts Institute of Technology, United States
Roberto Armenise	Poste Italiane, Italy
Robiah Yusof	Universiti Teknikal Malaysia Melaka, Malaysia
Roman Neruda	Institute of Computer Science, Czech Republic
S. Ramakrishnan	Dr. Mahalingam College of Engineering and Technology, India
Sabrina Ahmad	Universiti Teknikal Malaysia Melaka, Malaysia
Sadaaki Miyamoto	University of Tsukuba, Japan
Santi Llobet	Universitat Oberta de Catalunya, Spain
Satrya Fajri Pratama	Universiti Teknikal Malaysia Melaka, Malaysia
Saurav Karmakar	Georgia State University, United States
Sazalinsyah Razali	Universiti Teknikal Malaysia Melaka, Malaysia
Sebastian Ventura	University of Cordoba, Spain
Selva Rivera	Institute of C.E.D.I.A.C, Argentina
Shang-Ming Zhou	University of Wales Swansea, United Kingdom
Shyue-Liang Wang	National University of Kaohsiung, Taiwan
Siby Abraham	University of Mumbai, India
Silvia Poles	EnginSoft, Italy
Silvio Bortoleto	Federal University of Rio de Janeiro, Brazil
Siti Rahayu Selamat	Universiti Teknikal Malaysia Melaka, Malaysia
Steven Guan	Xi'an Jiaotong-Liverpool University, China
Sung-Bae Cho	Yonsei University, Korea
Swati V. Chande	International School of Informatics and Management, India
Sylvain Piechowiak	Universite de Valenciennes et du Hainaut-Cambresis, France
Takashi Hasuike	Osaka University, Japan
Teresa Ludermir	Federal University of Pernambuco, Brazil
Thomas Hanne	University of Applied Sciences Northwestern Switzerland, Switzerland
Tsu-Yang Wu	Harbin Institute of Technology Shenzhen Graduate School, China
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Vaclav Snasel	VŠB Technical University of Ostrava, Czech Republic
Valentina Colla	Scuola Superiore Sant'Anna, Italy
Victor Hugo Menendez Dominguez	Universidad Autonoma de Yucatan, Mexico

Vincenzo Loia	University of Salerno, Italy
Vincenzo Piuri	University of Milan, Italy
Virgilijus Sakalauskas	Vilnius University, Lithuania
Vivek Deshpande	MIT College of Engineering, India
Vladimir Filipovic	University of Belgrade, Serbia
Wei Wei	Xi'an University of Technology, China
Wei-Chiang Hong	Oriental Institute of Technology, Taiwan
Wen-Yang Lin	National University of Kaohsiung, Taiwan
Wilfried Elmenreich	University of Klagenfurt, Austria
Yasuo Kudo	Muroran Institute of Technology, Japan
Ying-Ping Chen	National Chiao Tung University, Taiwan
Yun-Huoy Choo	Universiti Teknikal Malaysia Melaka, Malaysia
Yunyi Yan	Xidian University, China
Yusuke Nojima	Osaka Prefecture University, Japan

Contents

Part VII: Computer Networks and Mobile Computing

Part VIII: Multimedia Signal Processing and Classification

Part IX: Intelligent Data Analysis and System Reliability Modeling

Part I

Innovative Computing Technology for Applications

High Availability and High Scalability to in-Cloud Enterprise Resource Planning System

Bao Rong Chang[1,*], Hsiu-Fen Tsai[2], Ju-Chun Cheng[1], and Yun-Che Tsai[1]

[1] Department of Computer Science and Information Engineering,
National University of Kaohsiung 81148, Taiwan
{brchang,l0985504,m1025512}@nuk.edu.tw
[2] Department of Marketing Management, Shu Te University, Kaohsiung 82445, Taiwan
soenfen@stu.edu.tw

Abstract. In this paper the host system architecture with high availability and high scalability has introduced to an in-cloud Enterprise Resources Planning (in-cloud ERP) deployed in the virtual environment to tackle the crucial problem of unexpected down-time or the failure of system failover that causes data loss and system terminated. Access control authentication has been adopted in the cloud to prevent the service-oriented hosts form external fraud or intrusion. As a result, the experiments have shown that the number of access for in-cloud ERP is 5.2 times as many as in-house ERP. The total cost of in-cloud ERP has decreased significantly to 48.4% of total cost of in-house ERP. In terms of operational speed, the approach proposed in this paper outperforms significantly two well-known benchmark ERP systems, in-house ECC 6.0 and in-cloud ByDesign.

Keywords: in-Cloud Enterprise Resources Planning (in-Cloud ERP), High Availability, High Scalability, Biometrics, Fraud.

1 Introduction

In this paper we introduce high availability and high scalability architecture for an in-cloud services solution to tickle the crucial problem of unexpected down-time or system failure to prevent data loss and system termination, as well as make good use of virtual machine cluster [1][2][3] approach to resolve the failover problem. This paper introduces in-cloud Enterprise Resources Planning (ERP) [4][5] in virtual environment and mobile devices users can easily access the in-cloud services via wired or wireless network with access control authentication [6]. As shown in Fig. 1, a open source ERP, OpenERP [7], has deployed successfully. Additionally, its access control authentication [8] [9] has brought into the virtual machine to achieve identity verification, safe sign-in, and attendance audit, as shown in Figs. 2 and 3. Then, detecting potential BotNet [10] and malicious attacks [11] in the network can efficiently increase the network security.

* Corresponding author.

J.-S. Pan et al. (eds.), *Intelligent Data Analysis and Its Applications*, *Volume 2*,
Advances in Intelligent Systems and Computing 298,
DOI: 10.1007/978-3-319-07773-4_1,

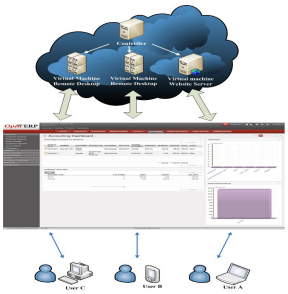


Fig. 1. OpenERP deployment

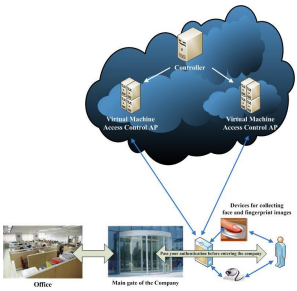


Fig. 2. Access control in a firm

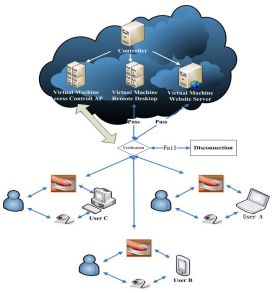


Fig. 3. Access control authentication in cloud

2 In-Cloud ERP and Access Control Authentication

Virtual machine clustering system in cloud is an integration of virtualization, virtual machines, and virtual services so that it can make existing resources be fully applied, such as VMware ESX/ESXi Server [12], Microsoft Hyper-V R2 [12] or Proxmox Virtual Environment [13]. This system can let users run many operating systems in a single physical computer simultaneously which largely decreases the expense of purchasing PCs. Most important of all, it has the following major functions, including virtual machine live migration, virtual storage live migration, distributed resource scheduling, high availability, fault tolerance, backup and disaster recovery, the transfer from physical machines to virtual machines, direct hardware accessing, virtual network switching, and so forth. This study introduces Proxmox Virtual Environment as the cloud computing and service platform with the virtual environment. The kernel-based virtual machine (KVM) acts as the main core of virtual machine, and it has installed the kernel of Linux-based operating system. OpenERP is adopted in this study as ERP application which provides many solutions for open sources software in the future, having it more expandable, making a great progress on cost deduction. The in-cloud ERP is established as follows:

1. Build Proxmox VE virtual machine cluster, and through WebPages manage the virtual machine. The webpage of login and management are shown individually in Fig. 4 and Fig. 5.
2. Create a virtual machine and set up its guest operating system in Proxmox VE virtual machine cluster.
3. Set up OpenERP in virtual machine, inclusive of OpenERP AP, PostgreSQL database, and web interface for end-user.
4. Sign-in at http://localhost:8096 or http://IP:8096 with the browser on virtual machine, pop up a login page of OpenERP as shown in Fig. 6, and then login to administrator to install the necessary modules as a result of an interface of user management as shown in Fig. 7.
5. Set up AP Server for biometric measures security [14]. When users sign-in, it will collect users' biometric features with capturing devices at client side as the evidence of legal or illegal sign-in [15].

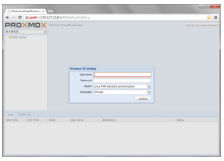


Fig. 4. Login to Proxmox VE Virtual machine cluster server

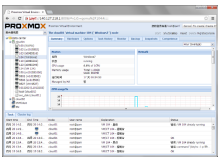


Fig. 5. Management web of virtual machine cluster server

Fig. 6. Remote login to OpenERP system

Fig. 7. User management in OpenERP

6. Collect the traffic flow in network and apply association rule analysis for establishing an identifiable and detectable system to automatically distinguish the network traffic flow either coming from potential BotNet or normal condition [16], as shown in Fig. 8.
7. Employ analysis of Multivariate Normal Model and Markov Chain to detect malicious attacks [17], as shown in Fig. 9.

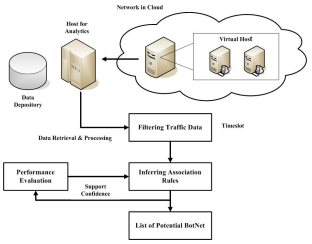


Fig. 8. Potential BotNet detection in cloud

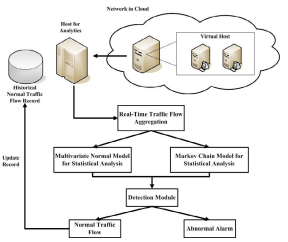


Fig. 9. Malicious attacks detection in cloud

3 High Availability and High Scalability Architecture

3.1 Virtual Machine High Availability

1. **Virtual Machine Live Migration:** when an execution error occurs at a node and causes an interruption, virtual machines at that node can be migrated themselves to the other nodes in which the left tasks of the failure node are also to be continued herein. A prerequisite is to ask for a shared storage as well as two units or more servers, for example, a Proxmox VE system as shown in Fig. 10.
2. **Virtual Storage Live Migration:** the system provides HA in virtual machines and accordingly HA will also support to virtual storage as well. Generally connecting a shared storage (e.g., SAN), the system may achieve the purpose of reaching a low downtime. When an execution error occurs at a node and causes an interruption, virtual storage at that node can be migrated itself to the other nodes to resume the left tasks of the failure node.
3. **Distributed Resource Scheduling:** Virtual machine management system such as Hyper-V [17] imports Non-uniform Memory Access (NUMA) mechanism for the

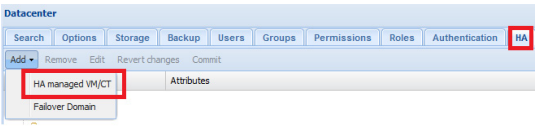


Fig. 10. HA optional setting of VM in Proxmox VE

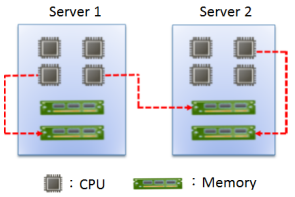


Fig. 11. Hardware resources allocation based on NUMA in Hyper-V R2

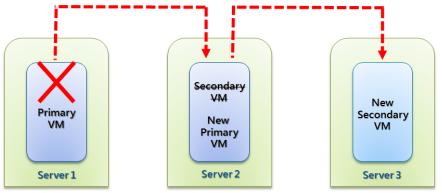


Fig. 12. Fault tolerance mechanism supported by VMware vSphere

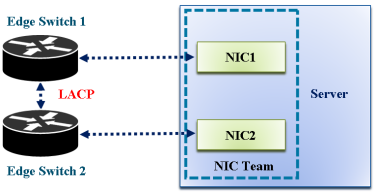


Fig. 13. Realizing the architecture of network high availability

resources allocation, in which computing cores and memory are divided into nodes, and each virtual machine attaches the corresponding node in accordance with the amount of the allocation of resources. That is, the resources of a virtual machine may be allocated from different server hardware resources as shown in Fig. 11.

4. **Fault Tolerance:** The main principle of reaching a zero downtime such as VMware vSphere [16] is that when a primary virtual machine is running, the system automatically generates a redundant virtual machine, totally equal to the primary one, located in other servers to synchronize the task. Once the system detects the primary virtual machine failure, the running task is immediately transferred to the redundant virtual machine, this redundant virtual machine becomes the primary virtual machine at once, and the system will replicate another redundant virtual machine once again as shown in Fig. 12.

3.2 Network High Availability

With Link Aggregation Control Protocol (LACP) [18], network interface cards can utilize Network Bounding techniques that will combine multiple network interface cards together, and in the meantime set the parameters of network interface card related to the HA function. For example, Linux systems can use the software ifenslave to gain fault-tolerant features in the combined network interface cards. That is, as one of network interface cards fails, work load will automatically switch to another one to carry on the successive networking tasks as shown in Fig. 13.

3.3 Storage High Availability

In general, storage device of iSCSI or NAS is able to provide hard drive array (RAID) function. If the system needs to consider both cost and performance, and fault tolerance solution, type of RAID 0+1 disk array is suggested to organize hard drive array. In addition, iSCSI or NAS storage device also probably risks the failure incident and hence the storage device needs to consider HA. At present, the storage device manufacturers have incorporated synchronous backup mechanism, but in contrary the traditional storage devices may not have this feature, which an additional server is required for implementing the synchronization between them as shown in Fig. 14.

According to HA of virtual machine, network, and storage as mentioned above, a diagram of in-cloud platform with high availability is illustrated in Fig. 15. With the minimum facility required for HA structure, the system needs at least two high-performance computing servers, two high-speed network switches, and two high-reliability storages to establish an in-cloud platform with HA.

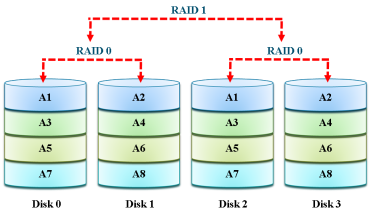


Fig. 14. RAID 0+1 system diagram

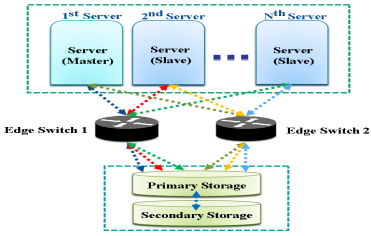


Fig. 15. an in-cloud platform with high availability

3.4 System High Scalability

In terms of the vertical scalability, virtualization technology can support high availability to ensure hardware maintenance and upgrade/downgrade, and when a single server requires a hardware upgrade/downgrade, virtual machines having the running services in a certain server can live migrate to the others without causing any interruption in these services. Similarly, in horizontal scalability, after virtual machines in a certain server have live migrated to the others, IT manager is able to remove the server from the virtual machine cluster or server cluster, not causing any service interruptions either. The SAN controller of these storage resources to realize so-called storage virtualization [19] could be integrated into different storage pools, which provides the established virtual volumes on demand for virtualized servers. Storage virtualization can more easily integrate heterogeneous storage devices so that changes in the storage devices have a higher elasticity and scalability. As shown in Figs. 16 and 17, they are illustrated to conduct scale-out and scale-in, respectively, in a cluster, and Figs. 18 and 19 have shown the scale-up and scale-down, respectively, of hardware specification in a node.

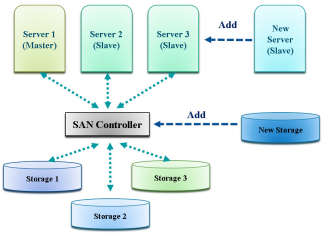


Fig. 16. Scale-out hardware facility

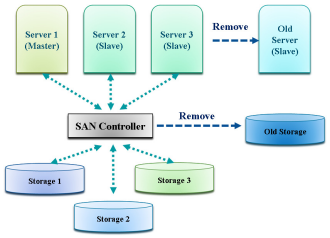


Fig. 17. Scale-in hardware facility

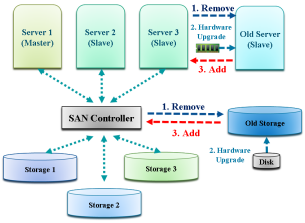


Fig. 18. Scale-up hardware specification

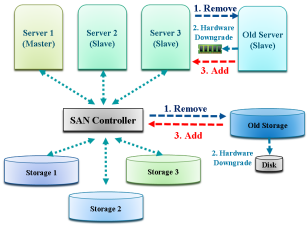


Fig. 19. Scale-down hardware specification

4 Experimental Results and Discussion

4.1 High Availability Testing

First in order to verify the high availability of the network, after the network used the function of Network Bonding, IT manager removed one of the network cables form an edge switch for a few seconds to check whether or not the network satisfies fault tolerance at this situation. After a test of ping command for 50 times, as a result, the connection quality is good because there is no packet loss during the cable removal, achieving the goal of network high availability as shown in Fig. 20. Next in order to verify whether the servers and storage devices achieve high availability, IT manager shut down a server on which a virtual machine was currently running, while the server-mounted storage device will correspondingly fail. Test results show that failover completed successfully because the virtual machine correctly transferred (migrated) to another server as shown in Fig. 21.

4.2 High Scalability Testing

In the virtual environment, the scalability is the most advantage for a cluster with several virtualized servers, and the changes in the scale of hardware devices occur quite often. Therefore the goal of this test in this section will emphasize the non-interrupted services when virtual machine migration is carried out. This test will use two servers denoted as cloud01 and cloud02, IT manager shutdowns server cloud02 normally where the virtual machines running in server cloud02 have migrated to server cloud01 automatically, and removes it from the cluster in order to simulate the

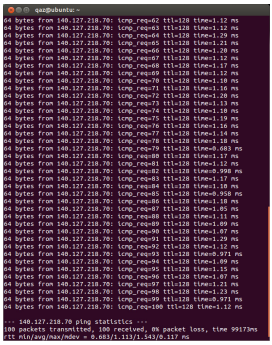
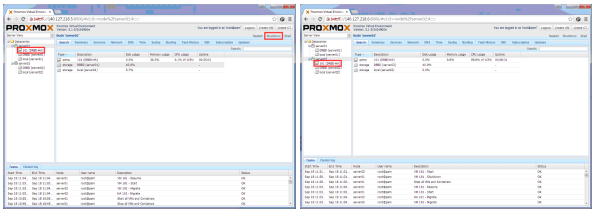


Fig. 20. Ping command to check the network quality after removing a network cable



(a) Before VM migration (b) After VM migration

Fig. 21. Failover by shutting down a server and initiating a virtual machine migration automatically

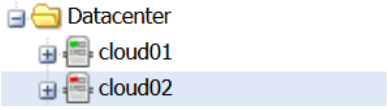


Fig. 22. Shutdown cloud02 in a cluster

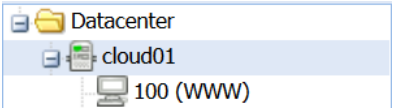


Fig. 23. Remove cloud02 in a cluster

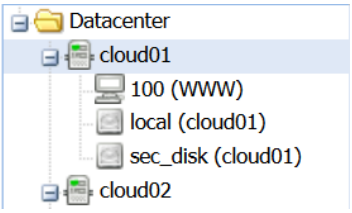


Fig. 24. Add cloud02 to a cluster

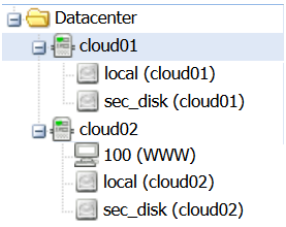
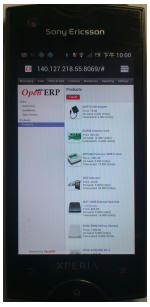


Fig. 25. A VM migrate to cloud02

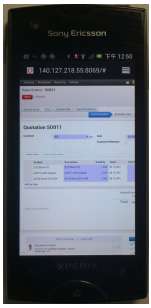
scale-down of a cluster as shown in Figs. 22 and 23. After that, IT manager re-boot server cloud02 normally and re-join it into the cluster, and IT manager conducts the virtual machines migrated back to server cloud02 from server cloud01 in order to simulate the scale-up of a cluster as shown in Fig. 24 and 25.

4.3 Access Control Authentication and ERP Testing

Users key-in at http://IP:8096 with the browser on Android smart phone to sign-in in-cloud ERP remotely via 3G/WiFi and next based on biometric measures the process of access control authentication is activated to capture human face and fingerprint at mobile device, deliver them to back-end server for identification, and then return the result back to mobile device. It takes about 2 seconds for identity verification as shown in Fig. 26. After that we begin to test ERP routines as shown in Fig. 27. Users

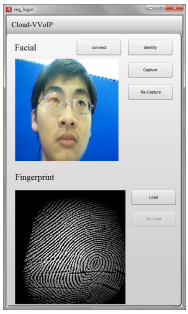


(a) List of products



(b) Sales order

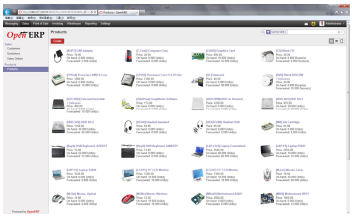
Fig. 26. Sign-in in-cloud OpenERP at smart phone



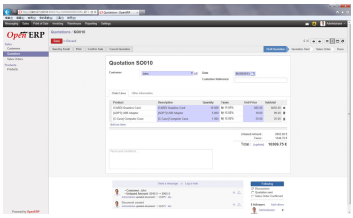
(a) capture the images

(b) identification

Fig. 27. Access control authentication using face recognition and fingerprint identification at smart phone



(a) List of products



(b) Sales order

Fig. 28. PC login to in-cloud OpenERP system

key-in at http://IP:8096 with the browser on personal computer to sign-in in-cloud ERP remotely via 3G/WiFi and then go for access control authentication at PC. After that we begin to test ERP routines as shown in Fig. 28.

4.4 Detection of Potential Botnet and Malicious Attacks

In cloud computing platform, Apriori algorithm based on association rules has been executed all the time. When invaders use a BotNet to link to another host or a single host location, the IP pairs between users to a BotNet and a BotNet to an attacked host will coincide more often, and thus the dependence between two IP pairs will be pretty high. As a result, the statistical outcomes derived from Apriori algorithm are fully collected and highlights potential BotNets from multiple period detection. According to the flow of malicious attacks and the relevant features of attack events, introducing the Multivariate Normal Model and Markov Chain is utilized to analyze two above-mentioned situations respectively and then developing a module with the ability of detection and identification is able to combine the two above statistical outcomes. When confronting different types of malicious attacks, it will be more efficient to identify possible causes, even finding the ways for immediate defense.

4.5 Assessment and Discussion

According to the experiments of online testing in the daily use of ERP in enterprise within a week, it was found that the growth rate of the use of in-cloud ERP increased dramatically approximate 5.2 times than the stand-alone ERP. In terms of the hardware cost in Taiwan, it costs the user $1,002.5 on the hardware equipment for a stand alone ERP, i.e. in-house ERP, in which the additional cost will be paid for air conditioning monthly fee of $18.4, space rent of $26.7, and hardware equipment maintenance fee of $16.7. In regard with the amortization expensive per month for a period of two years, the total expenditure costs $2,486.3. In other words, it costs an average monthly usage fee of $103.6. In contrast, renting an in-cloud ERP service in virtual environment only need about $50.1 monthly payment and it saves 1.07 times the cost of in-house ERP, i.e., reducing the total expenditure a lot. As shown in Table 1, a comparison of the number of accesses and the total expenditure for ERP, the proposed in-cloud ERP exclusively superior to in-house ERP. Two well-known benchmark ERP systems, ECC 6.0 [20] and ByDesign [21], are used to compare with the proposed one, according to ERP performance evaluation on system operational speed where the proposed approach outperforms the others as listed in Table 2.

Table 1. ERP assessment

Testing Item	Case A: in-house ERP	Case B: in-cloud ERP	Ratio of Case B to Case A
Number of Access (times/day)	63	328	5.206
Total Cost of Ownership (US dollars/month)	103.6	50.1	0.484

Table 2. Performance comparison of the operational speed of ERP (Unit: min-minutes, sec-seconds)

Function	ECC 6.0 (in-house ERP)	ByDesign (in-cloud ERP)	OpenERP (in-cloud ERP)
Create New Customer Master Data	7:10 min.	4:40 min.	3 min.
Create New Material Master	12:40 min.	10 min.	8:30 min.
Create Sales Order	5:20 min.	2 min.	1:30 min.
Search Function	2:10 min.	5 sec.	2 sec.
Average Access	6.83 min	4.19 min	3.26 min

5 Conclusion

This paper introduces high availability and high scalability architecture for an in-cloud ERP solution to tickle the crucial problem of unexpected down-time or system failure to prevent data loss and system termination, as well as make good use of virtual machine cluster approach to resolve the failover problem. As a result,

according to the experiments the proposed approach in this paper outperforms two well-known benchmark ERP systems, in-house ECC 6.0 and in-cloud ByDesign.

Acknowledgments. This work is fully supported by the National Science Council, Taiwan, Republic of China, under grant number: NSC 100-2221-E-390-011-MY3.

References

1. Beloglazov, A., Buyya, R.: Energy Efficient Allocation of Virtual Machines in Cloud Data Centers. In: 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 577–578 (2010)
2. Laurikainen, R., Laitinen, J., Lehtovuori, P., Nurminen, J.K.: Improving the Efficiency of Deploying Virtual Machines in a Cloud Environment. In: 2012 International Conference on Cloud and Service Computing, pp. 232–239 (2012)
3. Sotiriadis, S., Bessis, N., Xhafa, F., Antonopoulos, N.: Cloud Virtual Machine Scheduling: Modelling the Cloud Virtual Machine Instantiation. In: 2012 Sixth International Conference on Complex, Intelligent and Software Intensive Systems, pp. 233–240 (2012)
4. Yang, T.-S., Choi, J., Zheng, X., Sun, Y.-H., Ouyang, C.-S., Huang, Y.-X.: Research of Enterprise Resource Planning in a Specific Enterprise. In: 2006 IEEE International Conference on Systems, Man, and Cybernetics, pp. 418–422 (2006)
5. de Carvalho, R.A., Monnerat, R.M., Sun, Y.-H., Ouyang, C.-S., Huang, Y.-X.: Development Support Tools for Enterprise Resource Planning. IT Professional Magazine, 39–45 (2008)
6. Wu, H.-Q., Ding, Y., Winer, C., Yao, L.: Network Security for Virtual Machine in Cloud Computing. In: 2010 5th International Conference on Computer Sciences and Convergence Information Technology, pp. 18–21 (2010)
7. OpenERP, `http://v6.openerp.com/`
8. Zhao, J.-G., Liu, J.-C., Fan, J.-J., Di, J.-X.: The Security Research of Network Access Control System. In: 2010 First ACIS International Symposium on Cryptography and Network Security, Data Mining and Knowledge Discovery. E-Commerce & Its Applications and Embedded Systems, pp. 283–288 (2010)
9. Metz, C.: AAA protocols: Authentication, Authorization, and Accounting for the Internet. IEEE Internet Computing 3(6), 75–79 (1999)
10. Zhang, L.-F., Persaud, A.G., Johnson, A., Yong, G.: Detection of Stepping Stone Attack Under Delay and Chaff Perturbations. In: 25th Annual International Performance, Computing, and Communications Conference, pp. 256–266 (2006)
11. Yang, H.-Y., Xie, L.-X., Xie, F.: A New Approach to Network Anomaly Attack Detection. Fuzzy Systems and Knowledge Discovery 4, 317–321 (2008)
12. Chang, B.R., Tsai, H.-F., Chen, C.-M.: Evaluation of Virtual Machine Performance and Virtualized Consolidation Ratio in Cloud Computing System. Journal of Information Hiding and Multimedia Signal Processing 4(3), 192–200 (2013)
13. Chang, B.A., Tsai, H.-F., Chen, C.-M.: Empirical Analysis of Server Consolidation and Desktop Virtualization in Cloud Computing. Mathematical Problems in Engineering 2013, Article ID 947234, 11 pages (2013)
14. Wayman, J.L.: Biometrics in Identity Management Systems. IEEE Security & Privacy 6(2), 30–37 (2008)

15. Chang, B.R., Huang, C.-F., Tsai, H.-F., Lin, Z.-Y.: Rapid Access Control on Ubuntu Cloud Computing with Facial Recognition and Fingerprint Identification. Journal of Information Hiding and Multimedia Signal Processing 3(2), 176–190 (2012)
16. Hsiao, H.W., Fan, W.C.: Detecting step-stone with Network Traffic Mining Approach. In: The Fourth International Conference on Innovative Computing, Information and Control, pp. 1176–1179 (2009)
17. Hsiao, H.W., Lin, C.S., Chang, S.Y.: Constructing an ARP Attack Detection System with SNMP traffic data mining. In: The 11th International Conference on Electronic Commerce, pp. 341–345 (2009)
18. Link Aggregation Control Protocol, http://www.ieee802.org/3/ad/public/mar99/seaman_1_0399.pdf
19. Tate, J.: IBM System Storage SAN Volume Controller (IBM Redbooks). Vervante, New York (2006)
20. Doedt, M., Steffen, B.: Requirement-Driven Evaluation of Remote ERP-System Solutions: A Service-oriented Perspective. In: 34th IEEE Software Engineering Workshop, pp. 57–66 (2011)
21. Elragal, A., Kommos, M.E.: In-House versus In-Cloud ERP Systems: A Comparative Study. Journal of Enterprise Resource Planning Studies 2012, Article ID 659957, 13 pages (2012)

Prostate Tumor Identification in Ultrasound Images

Chuan-Yu Chang[1], Meng-Yu Tu[1], and Yuh-Shyan Tsai[2]

[1] Department of Computer Science and Information Engineering,
National Yunlin University of Science and Technology, Yunlin, Taiwan
[2] Department of Urology, National Cheng Kung University Hospital, Tainan, Taiwan
chuanyu@yuntech.edu.tw

Abstract. There are various medical imaging instruments used for diagnosing prostatic diseases. Ultrasound imaging is the most widely used tool in clinical diagnosis. Urologist outlines the prostate and diagnoses lesions based on his/her experiences. This diagnostic process is subjective and heuristic. Active contour model (ACM) has been successfully applied to outline the prostate contour. However, application of ACM in outlining the contour needs to give the initial contour points manually. In this paper, an automatic prostate tumor identification system is proposed. The sequential floating forward selection (SFFS) is applied to select significant features. A support vector machine (SVM) with radial basis kernel function is used for prostate tumor identification. Experimental results showed that the proposed method achieved higher accuracy than those of other methods.

Keywords: prostate tumor, feature selection, support vector machine.

1 Introduction

The prostate is a compound tubuloalveolar exocrine gland of the male reproductive system in most mammals. Prostate cancer has been the fifth most common malignancy and the seventh leading cause of cancer deaths in Taiwan. More than 3,000 men are diagnosed annually with prostate cancer, of which approximately 1000 die. Prostate cancer has become a common male urinary tract cancer.

Prostate is located in the pelvis, below the bladder and before the rectum. Most of the early prostate cancers have no symptoms, it's often discovered during routine health examination. Prostate specific antigen (PSA) blood test and digital rectal examination of the prostate are methods for prostate cancer screening. In order to detect early prostate cancer, physicians usually apply the ultrasound imaging to visualize prostate for diagnosis if either elevated PSA or abnormal Digital Rectal Examination (DRE). However, diagnosis by DRE is very subjective and highly depending on the experiences of urologist.

Various imaging modalities, such as magnetic resonance imaging (MRI), computerized tomography (CT), and ultrasound (US) imaging, are widely used in the diagnosis of various diseases with the assistance of medical image analysis techniques. In which, US imaging is inexpensive, non-invasive, and easy to use. Thus,

J.-S. Pan et al. (eds.), *Intelligent Data Analysis and Its Applications*, *Volume 2*,
Advances in Intelligent Systems and Computing 298,
DOI: 10.1007/978-3-319-07773-4_2,

ultrasound images become the most commonly used in clinical diagnosis, such as inspecting the prostate[4,6,7,8], breast tumor[3] and thyroid nodule[2].

Physician have outlined prostate region in ultrasound image when clinical diagnosis, and then determine whether a tumor within the region, outlined tumor area for tumor diagnosis. However, outlining the tumor area in ultrasound images is highly relied on physicians' experiences, it's inefficient and subjective. Therefore, we proposed a prostate tumor identification system in ultrasound images is proposed in this paper.

This paper is organized as follows: Section 2 presents the proposed prostate tumor identification method. Experimental results are presented in Section 3, where we compared with other method. Conclusions are drawn in Section 4.

2 Proposed Method

In this paper, an automatic SVM-based prostate tumor identification in ultrasound images is proposed. First, a preprocessing is adopted to reduce the influence of speckle noise and therefore enhance the contrast of prostate region. Second, a total of 236 features are extracted from each ROI. Third, the sequential floating forward selection (SFFS) is applied to select significant features. Finally, a support vector machine (SVM) with radial basis kernel function is used for prostate tumor identification. Figure 1 shows the flow diagram of the proposed method. Details of the procedures are described in the following subsections.

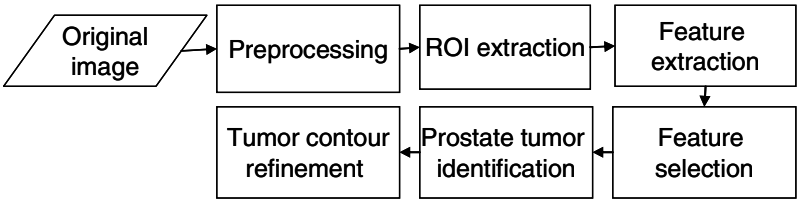
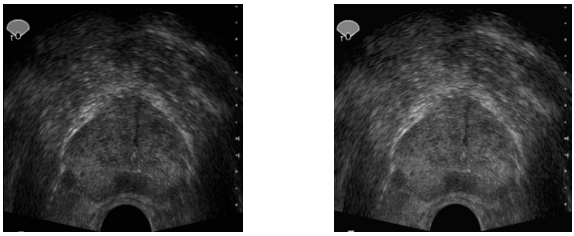


Fig. 1. Flow chart of the proposed method

2.1 Preprocessing

Since speckle noise and poor contrast in ultrasound images will affect the identification accuracy of the prostate tumor. Hence, speckle reduction by means of digital image processing should improve image quality and possibly the diagnostic potential of medical ultrasound.

In this paper, a 5×5 average filter [7] is used to reduce inevitable speckle noise in the prostate ultrasound images. A power-law [7] transformation is then applied to enhance the contrast; parameters c and r of power-law transformation are set as 1 and 0.7, respectively. Figure 2 shows the result of preprocessing. After preprocessing we can clearly see that the contrast between the prostate and background was strengthened.



(a) (b)

Fig. 2. (a) original image, (b) result after preprocessing

2.2 ROI Extraction

In order to identify the prostate tumor in the ultrasonic images, a lot of ROIs includes prostate tumor and non-tumor were outlined by an experienced urologist and confirmed by biopsy. Based on the sliding-windows method, a block with size of 23×23 is sliding on the outlined ROIs to segment patterns of tumor and non-tumor. The blocks are arrayed adjacent to each other with an $M\times M$ overlap. Figure 3 shows a case of outlined prostate tumor in ultrasound image. The extracted normal and tumor patterns are showed in Fig. 3(b) and (c), respectively.

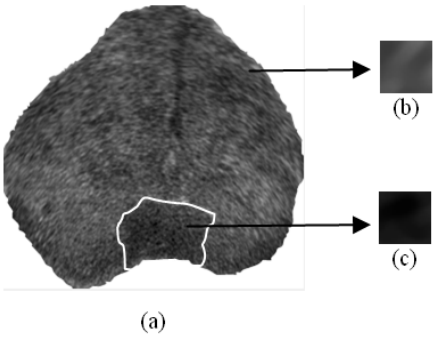


Fig. 3. (a) prostate region, (b) normal prostate pattern, (c) prostate tumor pattern

2.3 Feature Extraction

Textural features contain important information that is used for analysis and explanation of US images. In this paper, 236 features were extracted from the selected ROIs. Features are calculated from 11 different textural matrixes and transformations namely graylevel co-occurrence matrix (GLCM), Statistical Feature Matrix, Gray Level Run-Length Matrix, Laws' Texture Energy Measures, Neighboring Gray Level Dependence Matrix, Haar wavelet, homogeneity, histogram, block difference of inverse probabilities (BDIP) [10], Discrete Cosine Transform, and normalized multi-scale intensity difference (NMSID) [11].

2.4 Feature Selection

In previous stages, 236 features are obtained from each ROI. We can simply use them to identify prostate tumor, but it will be time-consuming. Feature selection is a common way to improve this situation. The sequential floating forward selection (SFFS) is adopted to sift significant features. The SFFS method consists of Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), which are capable of correcting wrong inclusions and removal decisions. Figure 4 shows the flow diagram of SFFS.

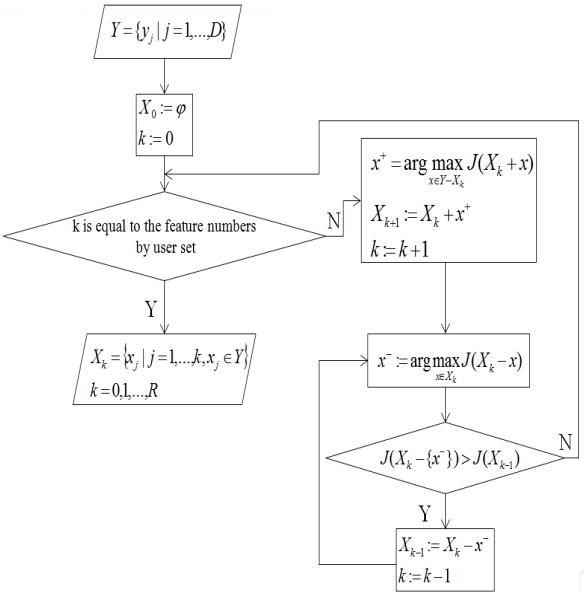


Fig. 4. The flow diagram of SFFS

In which, Y is features for all groups, D is the number of feature item, X is the best feature set selected from Y, R is the numbers of X, $J(x)$ is the cost function that evaluate the classification accuracy of feature set X. The terminative condition of the SFFS algorithm is the number of features k. The SFFS algorithm is summarized as follows:

Step 1: Input a feature group Y and set the terminative condition k.
Step 2: Use the function $J(x)$ to evaluated Y, that we can obtained a best feature x^+, then put the x^+ into the best feature group X.
Step 3: Use the function $J(x)$ to evaluated X to obtain a best feature x^-.
Step 4: Compared the best feature set without x^- ($J(X_k - x^-)$) and the best feature set of previous, if $J(X_k - x^-)$ is bigger than the best feature set of previous, then the best feature set of previous is equal to the best feature set without x^-, and k minus one then return to step 3. If $J(X_k - x^-)$ is not bigger than the best feature set of previous, then return to step 2 until reached the stop condition.

After the SFFS, five representative features were selected. Among them, four features are extracted from GLCM and one is extracted from textural energy matrix. The details of the selected features are described as follows:

The sum average of gray level co-occurrence matrix (GLCM) is defined as:

$$Sum\ Average = \sum_{i=2}^{2L} i p_{x+y}(i) \cdot \tag{1}$$

$$p_{x+y}(k) = \sum_{i=0}^{L} \sum_{\substack{j=0 \\ i+j=k}}^{L} p(i,j), \quad k = 2,3,...,2L \cdot \tag{2}$$

where L is the largest grey value in the image.

The contrast of gray level co-occurrence matrix (GLCM) is defined as:

$$Contrast = \sum_{n=0}^{L} n^2 \left(\sum_{i=0}^{L} \sum_{\substack{j=0 \\ |i-j|=n}}^{L} p(i,j) \right) \cdot \tag{3}$$

$p(i,j)$ is the values from gray level co-occurrence matrix after normalized, it defined as:

$$p(i,j) = \frac{C(i,j)}{\sum_{i,j} C(i,j)} \cdot \tag{4}$$

where $C(i,j)$ is the values of gray level co-occurrence matrix coordinate (i,j).

The difference entropy of gray level co-occurrence matrix(GLCM) is defined as:

$$Difference\ Entropy = -\sum_{i=0}^{L} p_{x-y}(i) \log\left(p_{x-y}(i)\right) \cdot \tag{5}$$

$$p_{x-y}(k) = \sum_{i=0}^{L} \sum_{\substack{j=0 \\ |i-j|=k}}^{L} p(i,j), \quad k = 0,1,...,L \cdot \tag{6}$$

The difference variance of gray level co-occurrence matrix (GLCM) is defined as:

$$DifferenceVariance = \sum_{i=0}^{L} \left(i - \sum_{j=0}^{L} j p_{x-y}(j) \right)^2 p_{x-y}(i) \cdot \tag{7}$$

In the present study, Law's texture energy [5] computed through the following two masks are used to obtain the texture features:

L5 (Level) = [1 4 6 4 1]

S5 (Spot) = [-1 0 2 0 -1]

The L5 vector gives a center-weighted local average. The S5 vector detects spots. In this study, a 2D convolution masks, namely S5L5 is obtained by computing the

outer products of pairs of 1D masks. By convoluting the image with the 2D mask, a texture energy maps are acquired. The statistics mean of the texture energy maps is then calculated and used as the texture feature.

2.5 Prostate Tumor Identification

Support vector machine (SVM) is a supervised classifier and it has been widely used in regression analysis and statistical classification. The basic idea of SVM is to map the input data into a high-dimensional feature space, and find the hyperplane that maximizes the margin between two classes. Fig. 5 shows the schematic of hyperplane. A SVM is used to identify the prostate tumor. In this paper, the SVM was implemented using the LIBSVM [1], and the radial basis function (RBF) defined as Eq.(8) was selected as the kernel function.

$$K(x^{(s)}, x_i) = \exp\left(\frac{-\left\|x^{(s)} - x_i\right\|^2}{2\sigma^2}\right). \tag{8}$$

where σ^2 is the width of the kernel.

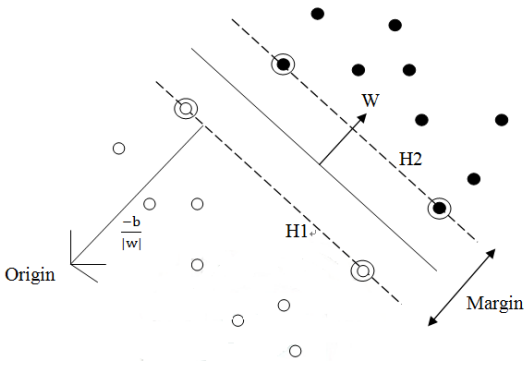


Fig. 5. Schematic concept of hyperplane

2.6 Tumor Contour Refinement

The above procedures can almost segment the prostate tumor completely. However, there still existed some artifact may be classified as part of prostate tumor. A morphological technique is utilized to separate the artifacts from a prostate tumor. Erosion with a 17×17 structuring element is derived to remove the connection between artifacts and prostate tumor. A region filling is then applied to fill holes in tumor region. Finally, dilation with a 17×17 structuring element is performed to recover the original size. The largest region is viewed as the tumor region.

3 Experimental Results

To show the capacity of the proposed method, the prostate ultrasound images for these experiments were taken at the urology department of National Chen Kung University Hospital. Resolution of the image is 3200×2400. A total of 35 cases were used in our experiments. In which, ten cases were chosen as training samples and the remaining cases were chosen as testing samples. The prostate tumor identification system was implemented by Microsoft Visual C# 2008 on a PC with 3.40 GHz Intel Core i7-2600 processor and 4GB RAM.

3.1 Preprocessing

The objective of the preprocessing is to reduce the influence of speckle noise and enhance the contrast of original ultrasound images. Figure 6 shows the results after preprocessing. The left column of Fig. 6 shows original ultrasound images. The results after the preprocessing were showed in the right column of Fig. 6. Obviously, after preprocessing we can clearly see that the contrast between the prostate and background was strengthened.

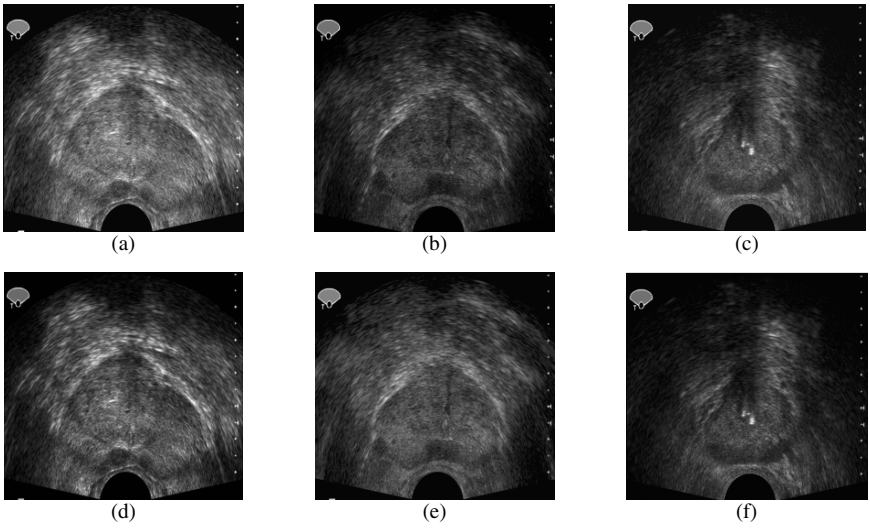


Fig. 6. (a-c) original prostate ultrasound image, (d-f) results after pre-processing

3.2 Prostate Segmentation and Tumor Classification

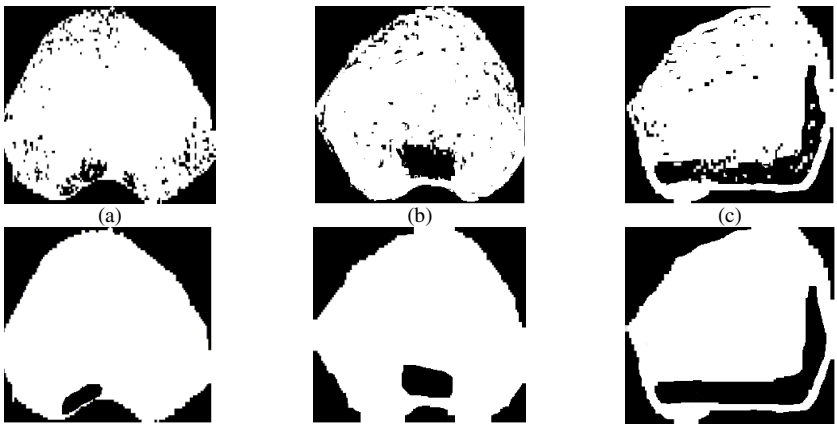
Before urologists diagnose prostate lesions, the contour of the prostate in US images must be manually outlined. However, manual segmentation is a time-consuming and non-reproducible task. Thus, an automatic prostate segmentation system combines the active contour model (ACM) is adopted [12]. The prostate classifier consists of a validation incremental neural network (VINN) and a radial-basis function neural network (RBFNN). Figure 7 shows the results of outlined the region of prostate in

ultrasound images. Obviously, the prostate region is correctly outlined. Figures 8(a, c, e) show the results of prostate tumor identification. There are many holes existed in prostates. Figures 8 (b, d, f) show the refinement results of Fig. 8 (a, c, e). The holes were removed, and prostate tumors were correctly outlined.

Figure 9 shows the prostate tumor identification results of the proposed method and outlined by urologist manually. Obviously, the outlined tumors of the proposed system are very close to the results outlined by experienced urologist.

(a) (b) (c)

Fig. 7. The outlined prostates in ultrasound images. (a) case1, (b) case2, (c) case3.



(a) (b) (c)

(d) (e) (f)

Fig. 8. (a-c) The results after prostate tumor identification. (d-f) The results after prostate tumor refinement.

3.3 Accuracy Evaluation of Prostate Tumor Segmentation

For a fair comparison, the accuracy was used to quantify the performance. Accuracy is defined as follow:

$$Accuracy = (N_{tp} + N_{tn}) / (N_p + N_n). \tag{9}$$

where N_p is the total number of pixels of the tumor, N_n is the total number of pixels of non-tumor region, N_{tp} is the number of actual prostate tumor detected by the proposed method, N_{tn} is the number of actual non-tumor pixels detected as

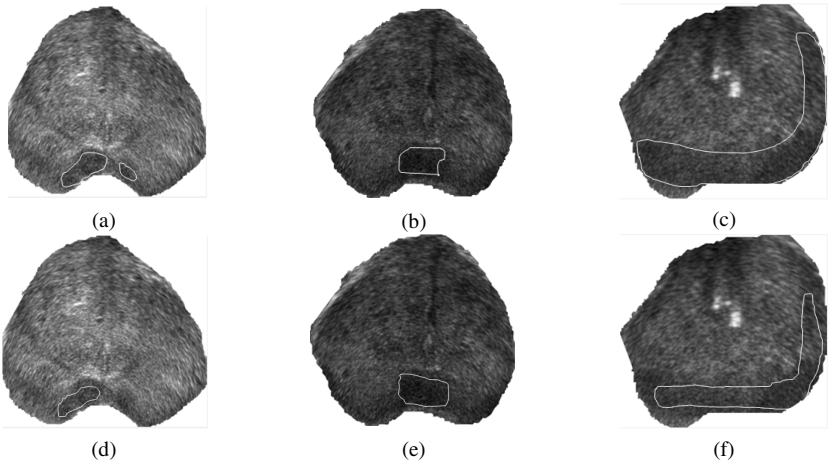


Fig. 9. (a-c)physician outlined tumor, (b-f) outlined tumors of the proposed method

non-tumor pixels, N_{fn} is the number of system classification as the non-tumor region in physician outlined tumor region, N_{fp} is the number of actual non-tumor pixels detected as tumor pixels.

To demonstrate the accuracy of the proposed method is higher than that of Zhang's method [9], Table 1 shows the comparison results. The accuracy of the proposed and Zhang's methods are 91.58% and 90.54%, respectively.

Table 1. Accuracy of Zhang's and proposed method

	The proposed method	Zhang's method
Accuracy	91.58%	90.54%

Table 2. The result for differential parameter

ROI size	Filter size	# of the best feature set	Accuracy
17×17	3×3	6	89.4%
17×17	5×5	5	87.67%
17×17	7×7	6	87.04%
19×19	3×3	7	89.08%
19×19	5×5	6	88.3%
19×19	7×7	6	87.06%
21×21	3×3	4	88.24%
21×21	5×5	5	89.49%
21×21	7×7	6	85.53%
23×23	3×3	4	91.56%
23×23	5×5	5	93.45%
23×23	7×7	6	91.85%
25×25	3×3	4	90.88%
25×25	5×5	5	91.74%
25×25	7×7	5	90.88%

3.4 Determination of ROI Size and Denosing Filter Size

To achieve a high accuracy, the size of ROI, masks, and the number of selected features should be appropriately set in the testing procedure. In order to obtain the appropriate parameters, various ROI sizes and denosing filter sizes were performed. Table 2 shows the accuracies for different parameters. The highest accuracy is achieved when the size of ROI and size of filter is set to 23×23 and 5×5, respectively.

4 Conclusion

In this paper, we proposed a SVM-based method for identification of prostate tumor in ultrasound images. The system can assist urologists on clinical diagnosis. The representative features after feature selection can save lots of time in classification and improve the classification results. According to the experimental results, the average accuracy rate has reach 91.58%. Our experiments have better accuracy than Zhang's method.

Acknowledgments. This work was supported in part by the by the National Science Council Taiwan, R.O.C, under the grants NSC 100-2628-E-224-003–MY3.

References

1. Chang, C., Lin, C.: LIBSVM: a library for support vector machines, 2.91. ed: Citeseer (2001)
2. Maroulis, D.E., Savelonas, M.A., Iakovidis, D.K., Karkanis, S.A., Dimitropoulos, N.: Variable Background Active Contour Model for Computer-Aided Delineation of Nodules in Thyroid Ultrasound Images. IEEE Trans. Infor. Tech. in Biomed. 11, 537–543 (2007)
3. Chen, D.R., Chang, R.F., Wu, W.J., Moon, W.K., Wu, W.L.: 3-D breast ultrasound segmentation using active contour model. Ultrasound in Medicine and Biology 29, 1017–1026 (2003)
4. Sahba, F., Tizhoosh, H.R., Salama, M.M.: Application of Reinforcement Learning for Segmentation of Transrectal Ultrasound Images. BMC Medical Imaging, 1471–2342 (2008)
5. Laws, K.: Textured image segmentation. University of southern california, Los Angeles Image Processing Institute (1980)
6. Betrouni, N., Vermandel, M., Pasquier, D., Maouche, S., Rousseau, J.: Segmentation of Abdominal Ultrasound Images of the Prostate Using a Priori Information and an Adapted Noise Filter. Computerized Medical Imaging and Graphics, 43–51 (2005)
7. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. (1992)
8. Mohamed, S.S., Salama, M.M.: Spectral Clustering for TRUS Images. BioMedical Engineering OnLine (2007)
9. Zhang, Y., Sankar, R., Qian, W.: Boundary delineation in transrectal ultrasound image for prostate cancer. Comp. Biol. Med., 1591–1599 (2007)
10. Chum, Y.D., Seo, S.Y.: Image Retrieval using BDIP and BVLC Moments. IEEE Transactions on Circuits and Systems for Video Tech. 13, 951–957 (2003)
11. Chen, E.L., Chung, P.C., Chen, C.L., Tsai, H.M., Chang, C.I.: An automatic diagnostic system for CT liver image classification. IEEE Trans. on Biomedical Engineering 45, 783–794 (1998)
12. Chang, C.Y., Tsai, Y.S., Wu, I.L.: Integrating Validation Incremental Neural Network and Radial-Basis Function Neural Network for Segmenting Prostate in Ultrasound Images. International Journal of Innovative Computing, Information and Control 7, 3035–3046 (2011)

QoS of Triple Play Services in LTE Networks

Lukas Sevcik, Karel Tomala, Jaroslav Frnda, and Miroslav Voznak

Department of Telecommunications, VSB-Technical University of Ostrava, 17. listopadu 15, 70833 Czech Republic
{lukas.sevcik.st1,karel.tomala,
jaroslav.frnda,miroslav.voznak}@vsb.cz

Abstract. This paper deals with studying the effects of performance LTE network throughput for data traffic. Utilisation rate of networks has a significant impact on the quality parameters of triple play services. LTE network was simulated by the software module additionally implemented in a development environment MATLAB. Throughput model has been obtained using this simulation that was used to test the QoS parameters for voice/video. Voice and video data streams using different codecs are transmitted for the obtained throughput. The measured qualitatively QoS parameters determine the resulting quality services from the perspective of end user perception.

Keywords: Quality of service, Long Term Evolution (LTE), E-model, Throughput, packet loss.

1 Introduction

Relative to increasing demands of mobile users for fast transmission of communications, enabling them to efficiently transmit both data and voice, but also multimedia documents, presentations or videos, the issue of QoS (Quality of Service) of multimedia services in modern data networks is topical theme. With incoming requests from users, directly proportional connected evolution in the development of new technologies for data transmission is through mobile networks from GPRS, through EDGE to contemporary 4G network using LTE technology. Currently, LTE technology is put in test and real traffic, so we decided to take a series of measurements, which showed features of this technology.

2 State of the Art

Network convergence refers to the provision of telephone, video and data communication services within a single network. For VoIP to be a realistic replacement for standard public switched telephone network (PSTN) telephony services, customers need to receive the same quality of voice transmission they receive with basic telephone services—meaning consistently high-quality voice transmissions. Like other real-time applications, VoIP is extremely bandwidth- and delay-sensitive. For VoIP transmissions to be intelligible to the receiver, voice packets should not be dropped,

J.-S. Pan et al. (eds.), *Intelligent Data Analysis and Its Applications*, *Volume 2*,
Advances in Intelligent Systems and Computing 298,
DOI: 10.1007/978-3-319-07773-4_3,

excessively delayed, or suffer various delay (otherwise known as jitter). Evaluation of the quality of transmitted voice incurred delays and packet loss is addressed in the work [2] a [3].

ITU-T G.107 recommendation [9] defines computation model - E-model, which takes into account all the links between transmission parameters. Its output is a scalar labelled R, which is a function of total expected call quality [1]. To evaluate speech quality, MOS (Mean Opinion Score) scale as defined by the ITU-T recommendation P.800 is applied [8].

The traditional method of determining voice quality is to conduct subjective tests with panels of human listeners. Extensive guidelines are given in ITU-T recommendations P.800/P.830. The results of these tests are averaged to give mean opinion scores (MOS), but such tests are expensive and impractical for testing in the field. Because of this, the ITU recently standardized a new model, Perceptual evaluation of Speech Quality (PESQ), that automatically predicts the quality scores that would be given in a typical subjective test. This is done by making an intrusive test and processing the test signals through PESQ [6].

Publication [10] deals with quality video and comparing the uploaded video codecs and their impact on the packet loss that causes errors in the transmission. In the work [7] we deal with focus on Packet loss and overall delay influence to objective QoS parameters of Triple play services. In the mentioned work, missing a comprehensive solution triple play services in real deployment; the impact of the policies used for the network and its quality parameters such as delay, jitter. The aim of our work and measurements in it was to analyze the quality of voice/video in various scenarios LTE network at the physical layer. From each scenario was used measured percentage network throughput. Testing the quality of voice/video was therefore focused on examining the impact of network utilization on its quality.

3 Methodology

3.1 Simulation throughput LTE

Technology LTE (Long Term Evolution) follows on HSPA and UMTS systems. It supports variable width of the frequency band ranging from 1.4 MHz to 20 MHz. The theoretical peak data rate in the downlink is 172.8 Mbit/s, in the uplink 57.6 Mbit/s. LTE system uses the principle of communication IP (MIP - Mobile IP). Communication thus takes place exclusively through the packet. The advantage is a low response in the radio interface (theoretically <10 ms) and high spectral efficiency [13] [14]. LTE system allows symmetrical (paired) or unsymmetrical (unpaired) communication. Symmetric communication using different uplink and downlink frequencies, which are mutually distant for the particular zone by a given value. The highest data rate can be achieved only when using symmetrical communication (FDD - Frequency duplex). When asymmetric communication is using to establish a connection to the constant frequency and communication takes place in alternating moments of time, so it is a time duplex (TDD).

In our case, LTE technology was simulated in an interactive environment MATLAB. We focused on the throughput parameter of this technology in certain scenarios set in the simulation of LTE. While simulating, we adjusted the frequency or time duplex (symmetric and asymmetric communication). Another parameter to measure was the ability to set the SNR, a signal-to-noise ratio. It is a parameter that indicates how many times the signal is fully reduced before we get to the noise level. It is expressed in decibels. Another parameter for testing and verification of the radio transmission, which affects throughput of LTE, were various options of the model channel.

3.2 Quality of Voice

During simulation of LTE, we implemented 4 types of codecs to compare quality of voice on the measured throughput of LTE. One type of codec was with two different bit rates. Capacity utilization of the network ensured IPERF tool. It can make capacity utilization to full at a given throughput LTE. We generated five voice traffic for each codec using program IxChariot with such busy network. Packets had not any priority, when voice was sending. Data were processed by FIFO. Parameters MOS, R-factor, the percentage packet loss and the percentage increase of jitter were evaluated after completion of voice traffic. These parameters evaluate toll IxChariot, which sets the parameters based on the E-model algorithm.

To measure quality of call, mathematical computer model, known as the E-model, was developed. It was designated specifically for the purpose of planning the transmission system, which takes into account many factors (particularly the impact of the delay on packet loss), that affect quality of voice in IP telephony. E-factor model assigns coefficient of deterioration to each partial factor that affects the quality of voice communication. Coefficients of deterioration are indicated in the following figure (figure.1), which presents connection between two participants. The output of the E-model is a value that is closer to the quality of the call in modern networks, either narrowband or broadband. It is called the R-factor and it is in the range of 0 to 100.

The R-factor is determined for the entire transmission chain. It takes into account not only the transmission channel, but also the end device. The higher the R-factor, the higher the quality of telephone service. The minimum acceptable value of R-factor is a ranking value 50. The resulting structure of mathematical relation of E-model is defined according to the recommendations by (1):

$$R = R_0 - I_s - I_d - I_{e-eff} + A \tag{1}$$

The R_0 represents the basic signal-to-noise ratio, which includes all kinds of noise, including noise caused by electrical circuits of equipment and noise caused on electrical wire.

$I_S$is the deterioration factor reflecting linear distortion of the transmission channel. This deterioration factor is the sum of a few individual factors, which may occur more or less simultaneously in conjunction with the transmission of the call.

Factor I_D represents all the deterioration which is caused by different combinations of delay.

I_{E-EFF} includes deterioration caused by the use of a voice codec, the manifestation of potential packet loss, and its resistance to loss. Finally, parameter A slightly improves the total final quality (for example in the case of satellite phone, the user is more focused to call than by normal use of solid terminal at home).

The final R-factor is a numerical value, usually between 15-94. With a value above 70, it is considered as acceptable. Parameter MOS is used as addition to R-factor. Generally known and most widely used is the five-point scale for quality rate. In this scale, it is possible to display display a value of R-factor (transfer described in recommendation G.107) [1].

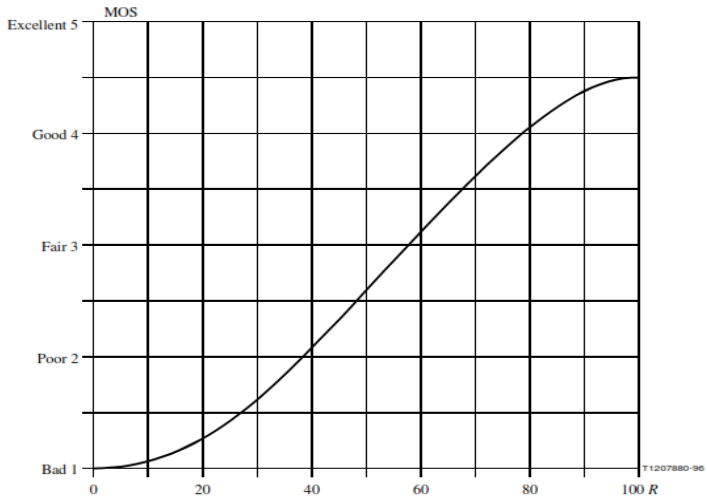


Fig. 1. R-factor as a function of the MOS

We tested four types of codecs in the measurement: G.711 A-law, G.729, G.726 a AMR.

G.711 is the most commonly used voice codec. It isnarrowband audio codec that provides the bit rate of 64 kbit/s. Transmitting audio signals in the range from 300 to 3400 Hz, the sampling rate is 8000 samples per second. A-law encoding takes a 13-bit linear sample as input and converts it into an 8 bit value.

G.729 is the most commonly used voice codec. It is narrowband audio codec that provides the bit rate of 64 kbit/s. Transmitting audio signals in the range from 300 to 3400 Hz, the sampling rate is 8000 samples per second. A-law encoding takes a 13-bit linear sample as input and converts it into an 8 bit value.

G.729 is very effective in comparison with the codec G.711 with relative to the final quality of the call. It uses a modified coding algorithm CELP, called CS-ACELP, bit rate of this codec is 8 kbit/s. Its disadvantage is the higher computational complexity. Total delay in the encoder is about value 25 ms.

G.726 is speech codec that uses the ADPCM coding involving the transmission of voice at rates of 16, 24, 32 and 40 kbit/s. It has size of sample 2, 3, 4 and 5 bits. The most commonly is use transmission rate 32 kbit/s, which doubles the usable capacity of the network by using half speed of G.711 codec.

3.3 Quality of Video

For measurement was selected two most used video formats for digital broadcasting, specifically MPEG-2 and MPEG-4 Part 10 (H.264). Both formats of video were in HD 720p resolution (1280x720), representing the standard of digital broadcasts at present time. Both formats were processed in a static image "Lenna".

Progam VLC was used to stream of both formats. Linux tool "iperf", simulating a UDP stream, was used for required network load. For the evaluation effects of degradation by network traffic to quality of video were chosen and used two objective methods for assessing the quality of video - PSNR and SSIM. These methods have been implemented by software MSU VQMT.

PSNR method is based on the mean square error (MSE), which is the square deviation between the test and the original sample. PSNR then determines the ratio between the maximum value of the signal to MSE in decibels. Basically, it can be said that while the MSE measures the difference between the two images, PSNR determines how closely the test image is similar to the original reference. Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as (2):

$$MSE = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[I(i,j) - K(i,j) \right]^2 \tag{2}$$

The PSNR is defined as (3):

$$\begin{aligned} PSNR &= 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \\ &= 20 \cdot \log_{10} \left(MAX_I \right) - 10 \cdot \log_{10} (MSE) \end{aligned} \tag{3}$$

where MAX is the maximum value that pixel can take (e.g. 255 for 8-bit image) [12].

SSIM method takes into account the perception of an image by the naked eye. It evaluate the visual impact of shifts the brightness of the image, change the contrast and other occurrence of errors in the transmitted picture (when compared it with the original image. Reference SSIM values are in the range 0-1, where zero means no similarity with the original and one is two completely identical pictures.

The resulting value of SSIM is a combination of three parameters, when for the original signal x and encoded signal y is valid the following equation (4) [7]:

$$SSIM(x,y) = [l(x,y)]\alpha[c(x,y)]\beta[s(x,y)]\gamma, \quad (4)$$

l(x,y) compares brightness of signal,
c(x,y) compares contrast of signal,
s(x,y) measures the correlation structure,
$\alpha > 0$, $\beta > 0$, $\gamma > 0$ controls the weighting of the various parts.

Diagram of the test platform is shown in the following figure:

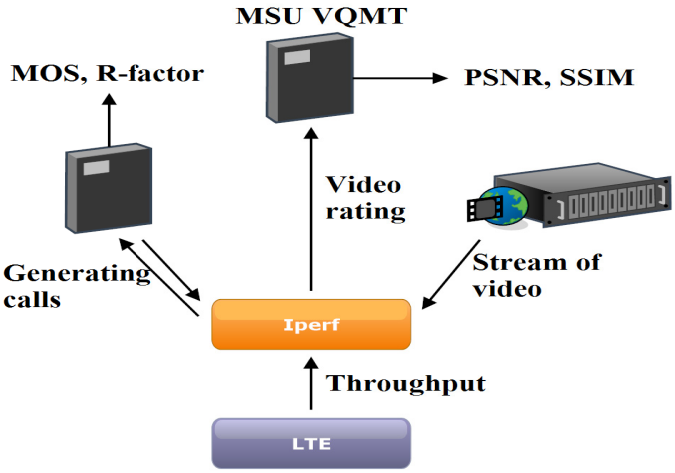


Fig. 2. Scheme of experiment

The simulation model were set through 4 channels:

- EPA (Extended Pedestrian A model): This is a multipath model of propagation of radio waves with pedestrian delay profile.
- EVA (Extended Vehicular A model): This is multipath model of propagation of radio waves with vehicular delay profile.
- ETU (Extended Typical Urban model): multipath model of propagation of radio waves with the town profile delays or wave reflections between buildings.
- HST (High Speed Train condition): It is a model of radio waves with a single component reflects. It represents the Doppler shift, which causes a result of high speed of train moving around the base station.

Some values obtained from tests and measurements are in the following reduced tables. The tables illustrate the relationship between parameters. The graph shows the change in signal-to-noise ratio of throughput at different LTE channel models. The values of signal to noise ratio (SNR) are suitably chosen so that it is possible to measure the throughput of between approximately 15 to 100% throughput.

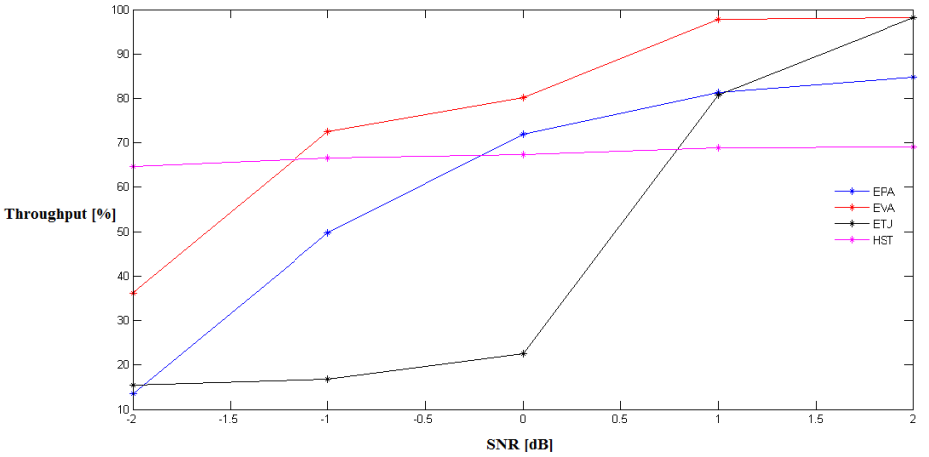


Fig. 3. The dependence of throughput to change SNR - time duplex

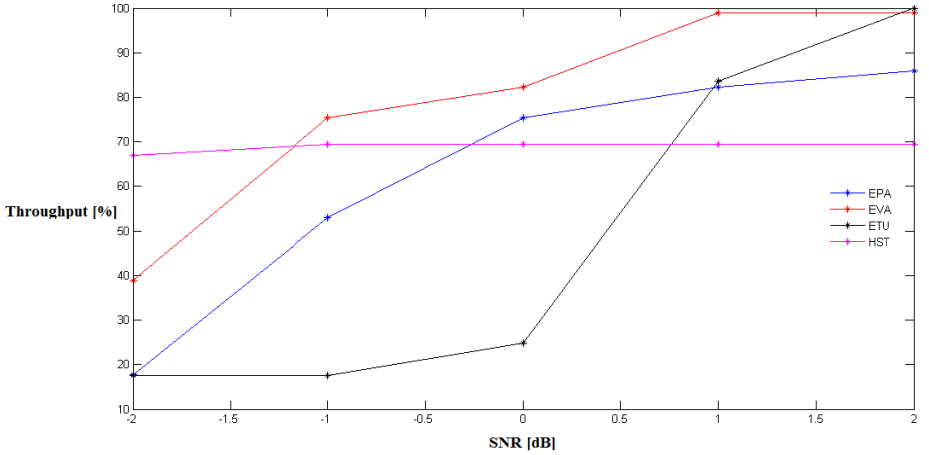


Fig. 4. The dependence of throughput to change SNR –frequencyduplex

Table 1. The quality of voice (codec G.711 A-law) by a throughput LTE with time duplex

Throughput	MOS	R-factor	Packetloss
13.51	3.012	58.616	0.182
22.56	3.362	66.694	1.9598
36.21	3.382	67.416	1.9732
49.73	3.508	69.994	1.4398
69.13	4.24	87.71	1.8534
81.25	4.33	90.22	1.3268
98.23	4.37	91.45	0.1

Table 2. The quality of video by a throughput LTE with time duplex

	MPEG-2			MPEG-4		
Throughput [%]	PSNR [db]	SSIM	PL [%]	PSNR [db]	SSIM	PL [%]
13,51	28,045	0,9396	3,48	17,048	0,7798	8,21
22,56	33,062	0,9533	0,83	19,291	0,8488	2,51
36,21	37,421	0,9557	0,81	21,254	0,8639	1,58
49,73	37,547	0,9562	0,78	21,404	0,8661	1,41
69,13	42,321	0,9658	0,12	24,962	0,9067	0,34
81,25	49,622	0,9827	0,07	38,152	0,9676	0,14
98,23	51,835	0.9897	0	51,2351	0,9848	0

4 Conclusion

Series of measurements were performed to test the characteristics of LTE technology and its impact on the quality of voice and video network. The difference between the time and frequency duplex in LTE had no significant imapact. Size of throughput was reflected in changing SNR and channel model. With increasing signal to noise ratio grew throughput, too. The results of the quality of voice show that at higher network bandwidth, the value of the R-factor increases and thereby the value of the MOS does too. The best results reported codec G.711 A-law at the highest throughput, respectively, it was the worst quality at the lowest throughput. Stable voice quality respectively, the difference between the worst and the best quality at the lowest and highest throughput recorded G.726 and AMR 12.2 kbit/s.

Results for quality of video with the decreasing throughput of the network show a decrease in quality of video and increase the packet loss in the network. Although H.264 provides fold higher compression ratio than the older MPEG-2, there is visible significant drop in quality with the decreasing throughput and packet loss in the network. Practical tests verify the basic hypothesis of lower throughput in the network that automatically decreases the quality of service. Sufficient QoS was achieved with throughput about 50%.

Acknowledgment. The research leading to these results received funding from the European Regional Development Fund in the IT4 Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Development of human resources in research and development of latest soft computing methods and their application in practice project (CZ.1.07/2.3.00/20.0072) funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic and partially was supported by the project SGS No. SP2014/72.

References

1. Voznak, M.: E-model modification for case of cascade codecs arrangement. International Journal of Mathematical Models and Methods in Applied Sciences 5(8), 1439–1447 (2011)

2. Managing Voice Quality with Cisco Voice Manager (CVM) and Telemate, http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_tech_note09186a00800946f8.shtml
3. Cole, R.G., Rosenbluth, J.H.: Voice over IP performance monitoring. ACM SIGCOMM Computer Communication (2001)
4. ITU-T P.800.1, Mean Opinion Score (MOS) terminology Geneva, ITU-T Recommendation P.800.1 (July 2006)
5. Karam, J.: A Comprehensive Approach for Speech Related Multimedia Applications. WSEAS Transactions on Signal Processing 6(1) (January 2010)
6. Rix, A.W., Beerends, J.G., Hollier, M.P., Hekstra, A.P.: Perceptual evaluation of Speech Quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs. In: IEEE Signal Processing Society International Conference on Acoustics, Speech, and Signal Processing, ICASSP (May 2001)
7. Frnda, J., Voznak, M., Rozhon, J., Mehic, M.: Prediction Model of QoS for Triple Play Services. In: 21st Telecommunications Forum TELFOR 2013 (2013)
8. ITU-T P.800, Methods for subjective determination of transmission quality, ITU-T Recommendation P.800, Geneva (August 1996)
9. ITU-T G.107, The E-model, a computational model for use in transmission planning, ITU-T Recommendation G.107, ITU-T Geneva (May 2010)
10. Feamster, N., Balakrishnan, H.: Packet Loss Recovery for Streaming Video. In: 12th International Packet Video Workshop, Pittsburgh, PA (2002)
11. Psytechnics Limited: PESQ: An Introduction, http://www.sageinst.com/downloads/960B/wp_pesq_introduction.pdf
12. Uhrina, M., Hlubik, J., Vaculik, M.: Correlation Between Objective and Subjective Methods Used for Video Quality Evaluation. Journal Advances in Electrical and Electronic Engineering 11(2) (December 2013) ISSN: 1804-3119
13. Dahlman, E., Parkvall, S., Sköld, J., Beming, P.: HSPA and LTE for Mobile Broadband. Elsevier (2007)
14. Parkvall, S., Dahlman, E., Furuskär, A., Jading, Y., Olsson, M., Wänstedt, S., Zangi, K.: LTE-Advanced – Evolving LTE towards IMT-Advanced. In: Vehicular Technology Conference 2008, pp. 1090–3038 (2008) ISSN: 1090-3038

New Genetic Algorithm for the p-Median Problem

Pavel Krömer[1,2] and Jan Platoš[2]

[1] Department of Computer and Electrical Engineering,
University of Alberta, Edmonton AB T6G 2V4, Canada
pavel.kromer@ualberta.ca
[2] IT4Innovations & Department of Computer Science
VŠB Technical University of Ostrava
Ostrava, Czech Republic
{pavel.kromer,jan.platos}@vsb.cz

Abstract. The p-median problem is a well-known combinatorial optimization problem with several possible formulations and many practical applications in areas such as operational research and planning. It has been also used as a testbed for heuristic and metaheuristic optimization algorithms. This work proposes a new genetic algorithm for the p-median problem and evaluates it in a series of computational experiments.

Keywords: genetic algorithm, p-median problem, experiments.

1 Introduction

The p-median problem is an NP-hard combinatorial optimization problem with interesting real-world applications in various areas including operational research, planning [4], and clustering [10]. The p-median problem (PMP) can be defined in terms of operational research [9].

Definition 1 (p-median problem). *For a set of m users U, set of n facilities L, and distance matrix $D^{m \times n}$ representing the cost of serving user i from location j find a subset $P \subseteq L$ of exactly p facilities so that the sum of minimal values in the rows of the column-reduced matrix will be minimized:*

$$\min_{P \in \{P \subseteq L \,:\, |P|=p\}} \left\{ \sum_{i \in U} \min_{j \in P} d_{ij} \right\} \tag{1}$$

Alternatively, the PMP can be defined as an integer programming problem or a graph problem. The PMP can be seen as a general combinatorial optimization task of finding an optimal fixed-length subset of p elements out of n, $p < n$. It has a number of real-world applications and it was also used as a testbed for heuristic and metaheuristic optimization algorithms [9].

Intuitively, the problem is challenging due to the large search space of $\binom{n}{k}$ possible solutions. Formally, it is known that the PMP is NP-hard [2,9].

J.-S. Pan et al. (eds.), *Intelligent Data Analysis and Its Applications, Volume 2,*
Advances in Intelligent Systems and Computing 298,
DOI: 10.1007/978-3-319-07773-4_4,

This study uses a novel pure metaheuristic genetic algorithm to solve the PMP and evaluates its performance in a series of computational experiments with PMP instances from the well-known OR-Library[1] of operational research problems.

The rest of this paper is organized in the following way: section 2 briefly introduces genetic algorithms as the metaheuristic optimization method used in this work to find good solutions to the p-median problem. Related work is summarized in section 3. The genetic algorithm proposed in this work is detailed in section 4 and numerical experiments are described in section 5. Section 6 concludes the study and outlines future work.

2 Genetic Algorithms

Genetic algorithms (GA) form a family of well known population-based metaheuristic soft optimization methods [8,1]. GAs solve complex optimization problems by the programmatic evolution of an encoded population of candidate problem solutions. The solutions are ranked using a problem specific fitness function. The artificial evolution is implemented by the iterative application of genetic operators and leads to the discovery of above average solutions. The basic workflow of a simple steady state GA [8] is shown in algorithm 1.

1 Define objective (fitness) function and problem encoding;
2 Encode initial population P of possible solutions as fixed length strings;
3 Evaluate chromosomes in initial population using objective function;
4 **while** *Termination criteria not satisfied* **do**
5 Apply selection operator to select parent chromosomes for reproduction: $sel(P_i) \rightarrow parent1$, $sel(P_i) \rightarrow parent2$;
6 Apply crossover operator on parents with respect to crossover probability to produce new chromosomes: $cross(pC, parent1, parent2) \rightarrow \{offspring1, offspring2\}$;
7 Apply mutation operator on offspring chromosomes with respect to mutation probability: $mut(pM, offspring1) \rightarrow offspring1$, $mut(pM, offspring2) \rightarrow offspring2$;
8 Create new population from current population and offspring chromosomes: $migrate(offspring1, offsprig2, P_i) \rightarrow P_{i+1}$;
9 **end**

Algorithm 1. A summary of the genetic algorithm

Problem encoding is an important part of the genetic search. It translates candidate solutions from the problem domain (phenotype) to the encoded search space (genotype) of the algorithm and defines the internal representation of

[1] `http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html`

the problem instances used during the optimization process. The representation specifies chromosome data structure and a decoding function [5]. The data structure defines the actual search space, its size and shape.

The crossover operator is the main operator of genetic algorithms distinguishing it from other population based stochastic search methods [8]. Crossover is primarily a creative force in the evolutionary search process. It is supposed to re-combine parent chromosomes in a stochastic manner and propagate building blocks (low order, low defining-length schemata with above average fitness) from one generation to another and to create new (higher order) building blocks by combining low order building blocks. It is intended to introduce to the population large changes with low disruption of the building blocks [12]. In contrast, mutation is expected to insert new material into the population by random perturbation of the chromosome structure. By doing this, however, new building blocks can be created or old ones disrupted [12].

Genetic algorithms have been successfully used to solve non-trivial multimodal optimization problems including data mining, classification, and prediction problems [13,11]. They inherit the robustness of emulated natural optimization processes and excel in browsing huge, potentially noisy problem domains. This study proposes a new genetic algorithm for the PMP. It defines suitable chromosome encoding, mutation and crossover operators and a fitness function to evaluate candidate PMP solutions.

3 Genetic Algorithms for the p-Median Problem

The p-median problem has been solved by a variety of genetic algorithms in the past [9].

A GA for the PMP is due to Correa et al. [4]. The authors developed a GA for the capacitated PMP in order to find good placement for university facilities with respect to the location and capacity of student residences. In capacitated PMP, each facility can serve only a limited demand. The algorithm used a set of facility indexes to represent selected subset of facilities and modified crossover and mutation operators to evolve the population of candidate solutions. The crossover constructed "exchange vectors" to indicate which alleles can be exchanged between parents. Mutation replaced one facility index by a randomly selected index of another facility that was not part of the solution yet. Additionally, the algorithm used a "heuristic hypermutation" operator that performed local search for good solutions. The proposed algorithm was compared to Tabu Search and the results suggested that the GA with heuristic hypermutation performs better than plain GA and Tabu Search with similar number of fitness function evaluations.

Alp et al. [2] proposed in 2003 another genetic algorithm for the PMP called ADE. The algorithm used specially tailored operators, steady state migration, complex chromosome initialization strategy, and a greedy heuristic to improve solutions found by the artificial evolution. Crossover operator was in ADE based on the union of parents' genes and a greedy method for deletion of extra genes.

No mutation operator was used in the algorithm since the authors reported it had no significant effect on the quality of solutions. The study showed that the proposed algorithm performed well and solved 28 uncapacitated PMP instances from the OR-Library to optimality.

A short paper by Lim and Xu [7] compared the performance of a fixed-length subset GA with and without problem specific heuristic recombination operators on the PMP. As in the previous cases, the GA used set representation of candidate solutions and additional heuristics to improve the solutions. It was concluded that the heuristics improve the quality of solutions found by this type of GA.

Pullan [10] used in 2008 a GA with cut-and-paste crossover operator and hybrid local search to find good PMP solutions. The local search was used to modify candidate solutions found by the evolution to have the required size and to explore the surrounding of evolved solutions. The experiments showed that the algorithm solved all 40 uncapacitated PMP instances from the OR-Library to optimality.

A multiobjective GA for the PMP is due to Arroyo et al. [3]. The algorithm looked for PMP solutions that would minimize both service costs and facility opening costs. Crossover operator of this GA used differences between parents to perform symmetric exchange of genes (whenever possible) and the mutation operator replaced randomly selected facility indexes by different indexes that were not part of the solution yet. The algorithm also used a path relinking strategy to explore search space between best individuals found by the GA in an attempt to improve the evolution.

A more recent work by Landa-Torres et al. [6] solved the capacitated PMP by a new grouping GA and grouping Harmony Search algorithm respectively. Both methods were improved by additional local search steps in order to improve their results. However, solution representation and genetic operators were quite different from those of the previous GAs due to the different nature of grouping algorithms.

This section reviewed recent GAs and GA-based methods for the PMP. A majority of the methods used local search and greedy steps in order to improve solution quality and most of them struggled with the generation of invalid solutions. Moreover, most reviewed works did not discuss computational aspects of chromosome decoding, genetic operator application, and local search steps. In this study we use a new simple GA for the PMP. It uses modified set encoding and only slightly altered crossover and mutation operators. The operations of the algorithm are, in contrast with some previous GAs for PMP, computationally inexpensive. The algorithm does not exploit any form of local search and does not use greedy steps in order to maintain generality. Therefore, it can be used for any other fixed-length subset selection problem when an appropriate fitness function is provided.

4 New Genetic Algorithm for the p-Median Problem

In order to design an efficient genetic algorithm for the PMP, we define the chromosome encoding, genetic operators, and fitness function. The encoding and operators are designed with the aim to obtain a compact representation, to exploit both crossover and mutation, and to avoid the creation of invalid individuals in the course of the evolution.

4.1 Chromosome Encoding

A valid subset of exactly k columns can be defined by the indices of the k columns selected from the full set of n columns in $A^{m \times n}$. No column can appear in the subset more than once in order to achieve the required subset size of k. The natural fixed-length encoding of a column subset is based on the indices of the selected columns, i.e. a chromosome $\boldsymbol{c}$ can be defined by eq. (2):

$$\boldsymbol{c} = (c_1, c_2, \ldots c_k), \quad \forall (i,j) \in \{0, \ldots n\} : c_i \neq c_j \tag{2}$$

However, such an encoding would be prone to the creation of invalid individuals in case traditional genetic operators such as the 1-point crossover or uniform mutation were applied. A few examples of the creation of invalid individuals from valid individuals are given in fig. 1 (conflicting genes are shown in **bold**). To avoid the creation of such invalid individuals, we slightly modify the encoding

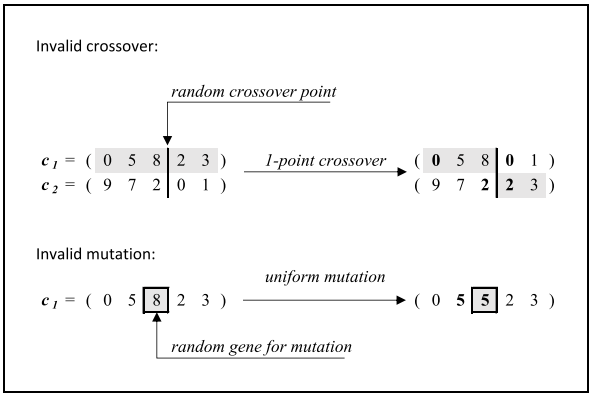


Fig. 1. Examples of the creation of invalid individuals

by sorting the indices within each chromosome, i.e. the encoding of a sample of k columns from n is defined by eq. (3):

$$\begin{aligned} &\boldsymbol{c} = (c_1, c_2, \ldots c_k), \\ &\forall (i,j) \in \{0, \ldots n\} : i < j \implies c_i < c_j \end{aligned} \tag{3}$$

With this modified encoding, the generation of random PMP individuals involves two steps. First, k unique columns out of all n possible columns are selected. Then, the indices in each individual are sorted in an increasing order. However, the ordering of the indices within the chromosome still requires the use of modified mutation and crossover operators that do not generate invalid individuals.

4.2 Genetic Operators

The genetic operators used by the proposed GA are based on the traditional mutation and crossover operators that were modified with respect to chromosome encoding so that they do not create invalid individuals and do not break the ordering of the chromosomes.

Order-Preserving Mutation. The order-preserving mutation operator replaces ith gene c_i in chromosome $\boldsymbol{c}$ with a value from the interval defined by its left and right neighbour as defined in eq. (4):

$$mut(c_i) = \begin{cases} urand^*(0, c_{i+1}), & \text{if } i = 0 \\ urand(c_{i-1}, c_{i+1}), & \text{if } i \in (0, n-1) \\ urand(c_{i-1}, N), & \text{if } i = N-1 \end{cases} \tag{4}$$

where $i \in \{0, \ldots, n\}$ and $urand(a, b)$ selects a uniform pseudo-random integer from the interval (a, b) (whereas $urand^*(a, b)$ selects a uniform pseudo-random integer from the interval $[a, b)$). Such mutation operator guarantees that the ordering of the indices within each chromosome remains valid. However, the order-preserving mutation of ith gene has no effect on chromosomes for which it holds that $(c_{i-1} + 1) = c_i = (c_{i+1} - 1)$.

Order-Preserving Crossover. The order-preserving crossover is based on the traditional one-point crossover operator [1,8]. It selects a random position i in parent chromosomes $\boldsymbol{c1}$ and $\boldsymbol{c2}$ and checks whether it is a suitable crossover point. A position i is a suitable crossover point if eq. (5) is true.

$$c1_i < c2_{i+1} \wedge c2_i < c1_{i+1} \tag{5}$$

If eq. (5) does not hold for i, the remaining positions in the chromosomes are scanned in the search for a suitable crossover point (i.e. first available position for which eq. (5) holds). It should be noted that an order-preserving crossover between 2 chromosomes might not always be possible. Examples of the use of the order-preserving crossover in different situations are shown in fig. 2.

4.3 Fitness Function

The fitness function used in this work is based on the definition of uncapacitated PMP in (1). Fitness of chromosome $\boldsymbol{c}$ is defined by eq. (6):

$$fit(\boldsymbol{c}) = \sum_{i \in U} \min_{j \in \boldsymbol{c}} d_{ij} \tag{6}$$

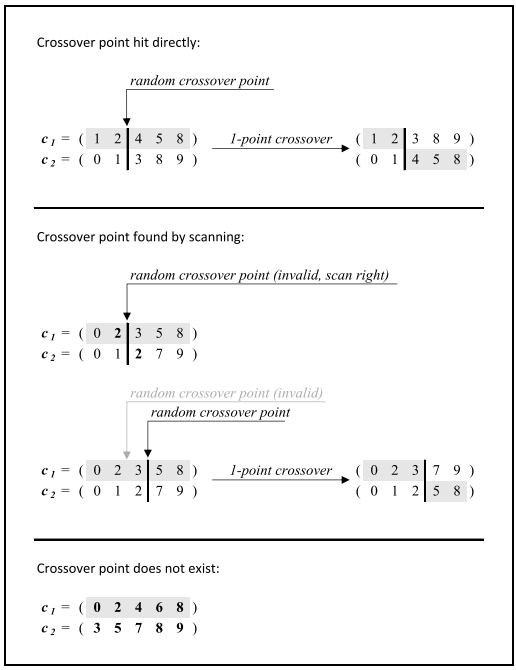


Fig. 2. Examples of order-preserving crossover

The GA for PMP minimizes $fit(\boldsymbol{c})$.

5 Experiments

In order to evaluate the proposed algorithm, a series of computational experiments was performed. The algorithm was implemented in C++ and used to find solutions to 40 uncapacitated PMP instances from the OR-Library. GA parameters used in the experiment were selected on the basis on initial trials and previous experience. The parameters were: steady state migration with generation gap 2 and elitism (only offspring with fitness better than worst member of current population are accepted), population size 100, crossover probability 0.8, mutation probability 0.4, and max. no. of generation 150000. Due to the stochastic nature of GA, all experiments were performed 30 times independently and presented results are averages over the 30 runs.

Overall, the average error of best solutions for all PMP instances found by the GA was 3.231%, the average error of average solutions was 5.46975% and the average error of worst solutions was 8.1845%. Results for different PMP instances are detailed in table 1. Names of the instances that were solved to optimality by at least one GA run are *emphasized*. It can be seen that the GA managed to find optimal solutions for 7 out of 40 instances in at least one run. This result is not as good as the results of previous genetic methods for PMP. However, it must be

Table 1. Service cost and error of solutions found for OR-Library PMP instances

PMP	Service cost found by GA			Solution error [%]		
Instance	best	worst	average (σ)	best	worst	avg.
pmed1	5819	5978	5884.133 (46.985)	0.00	2.73	1.12
pmed2	4105	4358	4202.500 (69.696)	0.29	6.47	2.68
pmed3	4254	4536	4347.500 (73.799)	0.09	6.73	2.29
pmed4	3063	3227	3148.967 (37.464)	0.96	6.36	3.79
pmed5	1392	1539	1468.200 (32.839)	2.73	13.58	8.35
pmed6	7824	7992	7911.233 (46.284)	0.00	2.15	1.11
pmed7	5642	5904	5729.267 (65.984)	0.20	4.85	1.75
pmed8	4565	4881	4687.100 (78.883)	2.70	9.81	5.45
pmed9	2871	3070	2966.100 (58.795)	5.01	12.29	8.49
pmed10	1364	1522	1441.767 (40.390)	8.69	21.27	14.88
pmed11	7702	7941	7783.267 (60.859)	0.08	3.18	1.13
pmed12	6683	7069	6767.667 (79.799)	0.74	6.56	2.01
pmed13	4555	4724	4623.733 (48.046)	4.14	8.00	5.71
pmed14	3151	3350	3248.433 (52.992)	6.17	12.87	9.45
pmed15	1858	2028	1962.800 (38.748)	7.46	17.29	13.52
pmed16	8162	8356	8235.400 (47.539)	0.00	2.38	0.90
pmed17	7037	7331	7169.333 (73.662)	0.54	4.74	2.43
pmed18	4944	5239	5051.800 (67.250)	2.81	8.94	5.05
pmed19	3020	3173	3088.500 (33.516)	6.15	11.53	8.56
pmed20	1982	2148	2061.267 (39.826)	10.79	20.07	15.22
pmed21	9192	9520	9338.767 (67.524)	0.59	4.18	2.20
pmed22	8653	8874	8755.367 (56.862)	0.86	3.44	2.06
pmed23	4797	5022	4903.833 (52.075)	3.85	8.72	6.17
pmed24	3155	3283	3216.367 (28.846)	6.55	10.87	8.62
pmed25	2028	2182	2112.933 (36.535)	10.94	19.37	15.59
pmed26	9917	10175	10045.467 (80.257)	0.00	2.60	1.30
pmed27	8343	8541	8428.100 (54.811)	0.43	2.82	1.46
pmed28	4636	4842	4755.100 (48.889)	3.07	7.65	5.72
pmed29	3253	3369	3322.600 (30.290)	7.25	11.08	9.55
pmed30	2231	2373	2280.200 (35.057)	12.17	19.31	14.64
pmed31	10086	10315	10184.400 (62.649)	0.00	2.27	0.98
pmed32	9394	9790	9488.233 (77.665)	1.04	5.30	2.06
pmed33	4910	5118	4986.800 (46.358)	4.47	8.89	6.10
pmed34	3250	3360	3311.533 (25.425)	7.87	11.52	9.91
pmed35	10400	10663	10527.233 (70.803)	0.00	2.53	1.22
pmed36	9980	10248	10121.700 (65.845)	0.46	3.16	1.89
pmed37	5317	5490	5390.533 (47.432)	5.14	8.56	6.60
pmed38	11060	11375	11171.900 (76.102)	0.00	2.85	1.01
pmed39	9471	9695	9584.933 (53.447)	0.51	2.89	1.72
pmed40	5358	5516	5440.667 (41.527)	4.49	7.57	6.10