Lecture Notes in Electrical Engineering 308

James J. (Jong Hyuk) Park
Shu-Ching Chen · Joon-Min Gil
Neil Y. Yen *Editors*

Multimedia and Ubiquitous Engineering

Springer

Lecture Notes in Electrical Engineering

Volume 308

For further volumes:
http://www.springer.com/series/7818

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

James J. (Jong Hyuk) Park · Shu-Ching Chen
Joon-Min Gil · Neil Y. Yen
Editors

Multimedia and Ubiquitous Engineering

Springer

Editors
James J. (Jong Hyuk) Park
Department of Computer Science and Engineering
Seoul National University of Science & Technology
Seoul
Korea, Republic of (South Korea)

Shu-Ching Chen
School of Computing and Information Sciences
Florida International University
Miami, Florida
USA

Joon-Min Gil
School of Information Technology Engineering
Catholic University of Daegu
Kyongsang-bukto
Korea, Republic of (South Korea)

Neil Y. Yen
School of Computer Science and Engineering
The University of Aizu
Aizu-Wakamatsu
Fukushima, Japan

ISSN 1876-1100 ISSN 1876-1119 (electronic)
ISBN 978-3-642-54899-4 ISBN 978-3-642-54900-7 (eBook)
DOI 10.1007/978-3-642-54900-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014936035

© Springer-Verlag Berlin Heidelberg 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

Message from the MUE 2014 General Chairs

MUE 2014 is the FTRA 8^{th} event of the series of international scientific conference. This conference takes place May 28–31, 2014, in Zhangjiajie, China. The aim of the MUE 2014 was to provide an international forum for scientific research in the technologies and application of Multimedia and Ubiquitous Engineering. It was organized by the Korea Information Technology Convergence Society in cooperation with Korea Information Processing Society. MUE2014 is the next event in a series of highly successful the International Conference on Multimedia and Ubiquitous Engineering, MUE-13 (Seoul, Korea, May 2013), MUE-12 (Madrid, Spain, July 2012), MUE-11 (Loutraki, Greece, June 2011), MUE-10 (Cebu, Philippines, August 2010), MUE-09 (Qingdao, China, June 2009), MUE-08 (Busan, Korea, April 2008), and MUE-07 (Seoul, Korea, April 2007).

The papers included in the proceedings cover the following topics: Multimedia Modeling and Processing, Ubiquitous and Pervasive Computing, Ubiquitous Networks and Mobile Communications, Intelligent Computing, Multimedia and Ubiquitous Computing Security, Multimedia and Ubiquitous Services, Multimedia Entertainment, IT and Multimedia Applications. Accepted and presented papers highlight new trends and challenges of Multimedia and Ubiquitous Engineering. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to Steering Chair: James J. (Jong Hyuk) Park (SeoulTech, Korea). Our special thanks go to the Program Chairs: Joon-Min Gil (Catholic University of Daegu, Korea), Neil Y. Yen (The University of Aizu, Japan), Shingchern You (National Taipei University of Technology, Taiwan), all Program Committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

May 2014

Jianhua Ma
Shu-Ching Chen
Timothy K. Shin
Doo-soon Park
Qingguo Zhou
MUE 2014 General Chairs

Message from the MUE2014 Program Chairs

Welcome to the FTRA 8th International Conference on Multimedia and Ubiquitous Engineering (MUE 2014), which will be held in Zhangjiajie, China on May 28–31, 2014. MUE 2014 will the most comprehensive conference focused on the various aspects of multimedia and ubiquitous engineering. MUE 2014 will provide an opportunity for academic and industry professionals to discuss recent progress in the area of multimedia and ubiquitous environment. In addition, the conference will publish high quality papers which are closely related to the various theories and practical applications in multimedia and ubiquitous engineering. Furthermore,we expect that the conference and its publications will be a trigger for further related research and technology improvements in these important subjects.

For MUE 2014, we received many paper submissions, after a rigorous peer review process, we accepted 40 articles with high quality for the MUE 2014 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation.We would like to sincerely thank the following invited speakers who kindly accepted our invitations, and, in this way, helped to meet the objectives of the conference: Prof. Han-Chieh Chao, National Ilan University, Taiwan and Prof. Timothy K. Shih, National Central University, Taiwan. Finally, we would like to thank all of you for your participation in our conference, and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

Joon-Min Gil, Korea
Neil Y. Yen, Japan
Shingchern You, Taiwan
MUE 2014 Program Chairs

Organization

Steering Chair

James J. Park	SeoulTech, Korea

General Chairs

Shu-Ching Chen	Florida International University, USA
Timothy K. Shih	National Central University, Taiwan
Doo-soon Park	SoonChunHyang University, Korea
Qingguo Zhou	Lanzhou University, China

Program Chairs

Joon-Min Gil	Catholic University of Daegu, Korea
Neil Y. Yen	The University of Aizu, Japan
Shingchern You	National Taipei University of Technology, Taiwan

Workshop Chairs

Namje Park	Jeju National University, Korea
Shijian Li	Zhejiang University, China
Michele Ruta	Politecnico di Bari, Italy
Vincent Huang	National Taichung University of Science and Technology, Taiwan
Chengcui Zhang	The University of Alabama at Birmingham, USA
KaLok Man	Xi'an Jiaotong-Liverpool University, China

Publication Chair

Hwa Young Jeong	Kyung Hee University, Korea

International Advisiory Committee

Young-Sik Jeong	Dongguk University, Korea
Makoto Takizawa	Seikei University, Japan
SeokCheon Park	Gachon University, Korea
Borko Furht	Florida Atlantic University, USA
Thomas Plagemann	University of Oslo, Norway
Roger Zimmermann	National University of Singapore, Singapore
Han-Chieh Chao	National Ilan University, Taiwan
Weijia Jia	City U. of Hong Kong, Hong Kong
Leonard Barolli	Fukuoka Institute of Technology, Japan
Hamid R. Arabnia	The University of Georgia, USA
Stephan Olariu	Old Dominion University, USA
Albert Zomaya	University of Sydney, Australia
Bin Hu	Lanzhou University, China
Yi Pan	Georgia State University, USA
Koji Nakano	University of Hiroshima, Japan
Qun Jin	Waseda University, Japan

Publicity Chairs

Kehua Guo	CSU, China
Mohamed Gaber	University of Portsmouth, UK
Jason C. Hung	Oversea Chinese University, Taiwan
Ryan Leong Hou U	Universidade De Macau, China
Chengjiu Yin	Kyushu University, Japan
Jian-Lian Chen	Aletheia University, Taiwan
Junbo Wang	University of Aizu, Japan
Bong-Hwa Hong	Kyung Hee Cyber University, Korea
Nan-Chen Hsieh	National Taipei University and Health Sciences, Taiwan
Deok-Gyu Lee	ETRI, Korea
Young-Ae Jeong	Sunmoon University, Korea

Local Arrangement Chairs

Kehua Guo	CSU, China
Ryan Leong Hou U	Universidade De Macau, China

Invited Speaker

Han-Chieh Chao	National Ilan University, Taiwan
Timothy K. Shih	National Central University, Taiwan

Program Committee

Akihiro Sugimoto	National Institute of Informatics, Japan
Angel D. Sappa	Universitat Autonoma de Barcelona, Spain
Bartosz Ziolko	AGH University of Science and Technology, Techmo, Poland
Ch. Z. Patrikakis	Technological Education Institute of Piraeus, Greece
Changsun Shin	Sunchon National University, Korea
Chantana Chantrapornchai	Silpakorn University, Thailand
Chao-Tung Yang	Tunghai University, Taiwan
Chi-Fa Chen	I-Shou University, Taiwan
Chih Cheng Hung	Southern Polytechnic State University, USA
DaeWon Lee	SeoKyeong University, Korea
Dakshina Ranjan Kisku	Asansol Engineering College, India
Dalton Lin	National Taipei University, Taiwan
Dariusz Frejlichowski	West Pomeranian University of Technology, Poland
Debzani Deb	Winston-Salem State University, USA
Dongkyun Kim	KISTI, Korea
Eung NamKo	Baekseok University, Korea
Ezendu Ariwa	London Metropolitan University, UK
Fay Huang	National Ilan University, Taiwan
Florian Stegmaier	University of Passau, Germany
Hae Young Lee	Seoul Women's University, Korea
Hai Jin	Huazhong University of Science and Technology, China
Hangzai Luo	East China Normal University, China
Hansung Lee	ETRI, Korea
Harald Kosch	University of Passau, Germany
Hari Om	Indian School of Mines University, India
HaRim Jung	Sungkyunkwan University, Korea
Helen Huang	University of Queensland, Australia
HeonChang Yu	Korea University, Korea
Hermann Hellwagner	Klagenfurt University, Austria
Hongsoo Kim	SK Planet, Korea
HwaMin Lee	Soonchunhyang University, Korea
Hyun-Wook Jin	Konkuk University, Korea
Jaehwa Chung	Korea National Open University, Korea
JongHyuk Lee	Samsung Electronics, Korea
Jong-Kook Kim	Korea University, Korea
Jungong Han	Civolution Technology, Netherlands
Jun-Won Ho	Seoul Women's University, Korea
Ki-Sik Kong	Namseoul University, Korea
Klaus Schoffmann	Klagenfurt University, Austria

Kwanghee Choi	LG Uplus, Korea
Kwang Sik Chung	Korea National Open University, Korea
KwonWoo Yang	Gongju National University of Education, Korea
Lukas Ruf	Consecom AG, Switzerland
Maria Vargas-Vera	Universidad Adolfo Ibanez, Chile
Mario Doeller	University of Applied Science, Germany
Maytham Safar	Kuwait University, Kuwait
Mihye Kim	Catholic University of Daegu, Korea
Ming Li	California State University, USA
Molina Helena	Universidad de Sevilla, Spain
Neungsoo Park	Konkuk University, Korea
Ning Zhou	University of North Carolina, USA
Pascal Lorenz	University of Haute Alsace, France
Quanqing Xu	Data Storage Institute, A*STAR, Singapore
Rachid Anane	Coventry University, UK
Ralf Klamma	RWTH Aachen University, Germany
Reinhard Klette	The University of Auckland, New Zealand
Sae-Hak Chun	SeoulTech, Korea
Sagarmay Deb	University of Southern Queensland, Australia
Sang-Seon Byun	Catholic University of Pusan, Korea
Savvas Chatzichristofis	Democritus University of Thrace, Greece
Seunghae Kim	KISTI, Korea
Shin Bok-Suk	The University of Auckland, New Zealand
Sokratis Katsikas	University of Piraeus, Greece
Soonyoung Jung	Korea University, Korea
Sung-Hwa Hong	Mokpo National Maritime University, Korea
SungSuk Kim	Seokyeong University, Korea
Teng Li	Anhui University, China
Thomas Grill	University of Salzburg, Austria
Toshihiro Yamauchi	Okayama University, Japan
Ui-Sung Song	Busan National University of Education, Korea
Wei Wei	Xi'an University of Technology, China
Weifeng Chen	California University of Pennsylvania, USA
Xubo Song	Oregon Health and Science University, USA
Yang Yang	National University of Singapore, Singapore
Yong-Hee Jeon	Catholic University of Daegu, Korea
Yong-Yoon Cho	Sunchon University, Korea
Young-Gab Kim	Catholic University of Daegu, Korea
Yunhee Kang	Baekseok University, Korea
Zheng-Jun Zha	National University of Singapore, Singapore

Message from AIAE2014 Workshop Chair

Welcome to the Sixth International Workshop on Artificial Intelligence Application on E-service (AIAE 2014).

Service-Oriented age is coming. User could achieve more and more emerging eservice applications and service via various digital carriers. To realize user behaviors of service-oriented age is becoming an important issue. A lot of researches had been successfully applied to artificial intelligence (AI) techniques in related research area, such as recommender system, search engine, information retrieve, etc. Thus, we completely accept that drawing attention to applied AI on the topic of e-service is an area that shows great research potential. It could be an important time to shift the research focus to the application area. The International Workshop on AIAE2014 brings together scientists, engineers, computer users, and students to exchange and share their experiences, new ideas, and research results about all aspects (theory, applications and tools) of computer and information science, and discuss the practical challenges encountered and the solutions adopted. The workshop on 'artificial intelligence application' aims to serve as an international forum for researchers and practitioners willing to present their early research results and share experiences in the field of future science and computer.

AIAE-2014 contains high quality research papers submitted by researchers from all over the world. Each submitted paper was peer-reviewed by reviewers who are experts in the subject area of the paper. Based on the review results, the Program Committee accepted 8 papers.

For organizing an International workshop, the support and help of many people is needed. We would like to thank all authors for their work and presentation, all members of the program committee and reviewers for their cooperation and time spent in the reviewing process. Finally, Special thanks are extended to the staffs from MUE 2014, who contributed so much to the success of the workshop.

AIAE-2014 Workshop Program Chairs
Chen-Shu Wang

AIAE-2014 Organization

General Chairs

Chen-Shu Wang	National Taipei University of Technology, Taiwan
Mei-Yu Wu	Chung Hua University, Taiwan
Vincent Tien-Chi Huang	National Taichung University of Science and Technology, Taiwan

Program Chairs

Chih-Kun Ke	National Taichung University of Science and Technology, Taiwan

Program Committee

Chi-Chung Lee	Chung Hua University, Taiwan
Yuan-Chu Hwang	National United University, Taiwan
Ming-Hsiung Ying	Chung Hua University, Taiwan
I-Hsien Ting	National University of Kaohsiung, Taiwan
Duen-Ren Liu	National Chiao Tung University, Taiwan

Message from IoV2014 Workshop Chair

Internet of Vehicles (IoV) is expected to become the core building blocks of the intelligent transportation systems (ITS) where encompass a broad range of wireless communications and electronics technologies. In order to offer new solutions to the driving safety, IoVs (including automobiles and the high-speed rail trains) involve mobile communication system of the vehicles over a public Internet to increase situational awareness and avoid collision risks through vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) data transmission. However, inherited from its own unique features (such as mobility pattern and ad hoc nature), on-going researches aim at improvement of spectrum efficiency, coverage, packet loss, mobility management, security and privacy, which are actively studied and developed in both wireless academia and industry. Cloning the same strategy of the Open Handset Alliance, Google teams up with auto industry leaders to form the Open Automotive Alliance (OAA) bringing the Android platform to the car which makes driving safer, easier and more enjoyable. On the other hands, Apple launches iOS in the Car (iOSitC) standards to enable iOS devices (e.g., iPhone and iPad) and its applications (e.g., Siri) interact with manufacturers' built-in in-car systems. IoV provides entrepreneurs with a stream of opportunities to develop new products and services. Therefore, this workshop serves to unite common research interests in IoV technologies to discuss related issues and share novel solutions.

We would like to thank the program committee of IoV2014 and the external reviewers for their constant support. We are very pleased to invite you to attend this workshop. Enjoy the program and your stay in Zhangjiajie.

Bo-Chao Cheng
National Chung-Cheng University, Taiwan

IoV-2014 Organization

Steering Committees

Chung-Ming Huang	NCKU, Taiwan
Xiang Cheng	Peking University, China
Bo Ai	Northern Jiaotong University, China
Phone Lin	National Taiwan University, Taiwan
Tao Zhang	Cisco System, USA

General Chairs

James J. (Jong Hyuk) Park	SeoulTech, Korea
Yang Xiao	University of Alabama, USA
Han-Chieh Chao	National Ilan University, Taiwan

Program Chairs

Bo-Chao Cheng	National Chung-Cheng University, Taiwan
Yuh-Shyan Chen	National Taipei University, Taiwan
Xiaohong Peng	Aston University, UK

Publicity Chairs

Huan Chen	National Chung-Hisng Univerity, Taiwan
Rafael Falcon	Larus Technologies, Canada
Xu Shao	Institute for Infocomm Research, Singapore

Program Committee

T. Russell Hsing	National Chiao Tung University, Taiwan
Zhikui Chen	Dalian University of Technology, China
Shouzhi Xu	China Three Gorges University, China

Tsang-Ling Shen	National Sun Yat-Sen University, Taiwan
Ibrahim Korpeoglu	Bilkent University, Turkey
Steven Lee	National Chung-Cheng University, Taiwan
Chien-Chung Shen	University of Delaware, USA
Jiunn-Liang Lin	ITRI, Taiwan
James C.N. Yang	National Dong Hwa University, Taiwan
Kwang-Hui Lee	Changwon National University, Korea
Jung-Shian Li	National Cheng-Kung University, Taiwan
Hui-Kai Su	National Formosa University, Taiwan

Message from DAES2014 Workshop Chair

Welcome to the 2014 International Workshop on Design and Analysis of Embedded Systems (DAES-14).

The field of design, synthesis, implementation, test and analysis of embedded systems including digital, analog and mixed devices is grown rapidly. The aim of the workshop is to bring together scientists, practitioners and students from universities and industry and provide them platform for presentations, discussions and experiences exchange in theory, applications and technology of this field. DAES'14 contains selected papers submitted by researchers from Europe, China and USA. Each submitted paper was peer-reviewed by experts in its subject area. Based on the review score, 8 papers were accepted for presentation.

We are grateful to all members of the program committee, who have done a remarkable job in reviewing papers submitted for the workshop. We also would like to thank all authors for their work and presentation. The organization of the workshop was be impossible without a support from MUE 2014 chairs and FTRA representatives. Therefore, we thank steering chair of MUE 2014 – James J. Park and workshop chairs of MUE 2014 – Namje Park and KaLok Man.

DAES'14 Workshop Program Chair
Arkadiusz Bukowiec

DAES-2014 Organization

Program Chair

Arkadiusz Bukowiec	University of ZielonaGóra, Poland

Program Co-Chairs

KaLok Man	Xi'an Jiaotong-Liverpool University, China
Piotr Mróz	University of ZielonaGóra, Poland

Honorary Chair

Marian Adamski	University of ZielonaGóra, Poland

Program Committee

Grzegorz Borowik	Warsaw University of Technology, Poland
Anikó Costa	New University of Lisbon, Portugal
Michał Doligalski	University of ZielonaGóra, Poland
Luis Gomes	New University of Lisbon, Portugal
Tomasz Gratkowski	University of ZielonaGóra, Poland
Wolfgang Halang	University of Hagen, Germany
Vladimír Janoušek	Brno University of Technology, Czech Republic
Grzegorz Łabiak	University of ZielonaGóra, Poland
Wiesław Miczulski	State School of Higher Professional Education in Sulechów, Poland
Adam Milik	Silesian University of Technology, Poland

Joao Monteiro	University of Minho, Portugal
Ann Pławiak-Mowna	University of ZielonaGóra, Poland
Alfredo Rosado-Muňoz	University of Valencia, Spain
Henry Selvaraj	University of Nevada Las Vegas, United States
Bernd Steinbach	Freiberg University of Mining and Technology, Germany
Jacek Tkacz	University of ZielonaGóra, Poland
Michal Vopálenský	College of Polytechnics Jihlava, Czech Republic
Remigiusz Wiśniewski	University of ZielonaGóra, Poland

Message from DATICS2014 Workshop Chair

The International Workshop DATICS-MUE'14: Design, Analysis and Tools for Integrated Circuits and Systems in the 8th FTRA International Conference on Multimedia and Ubiquitous Engineering will take place in Zhangjiajie, China, May 29-31, 2014.

DATICS Workshops were initially created by a network of researchers and engineers both from academia and industry in the areas of Design, Analysis and Tools for Integrated Circuits and Systems. Recently, DATICS has been extended to the fields of Communication, Computer Science, Software Engineering and Information Technology.

The main target of DATICS-MUE'14 is to bring together software/hardware engineering researchers, computer scientists, practitioners and people from industry to exchange theories, ideas, techniques and experiences related to all aspects of DATICS.

The International Program Committee (IPC) of DATICS-MUE'14 consists of about 150 experts in the related fields of DATICS-MUE'14 both from academia and industry. DATICS-MUE'14 is partnered with CEOL: Centre for Efficiency-Oriented Languages (Ireland), Minteos (Italy), KATRI (Japan & Hong Kong), Distributed Thought (UK), Baltic Institute of Advanced Technology - BPTI (Lithuania), Solari (Hong Kong), Transcend Epoch (Hong Kong) and Xi'an Jiaotong-Liverpool University - XJTLU (China - UK).

The DATICS-MUE'14 Technical Program includes 6 papers which are organized into lecture sessions. On behalf of the IPC, we would like to welcome the delegates and their guests to the DATICS-MUE'14. We hope that you will enjoy the workshop and find the DATICS-MUE'14 Technical Program exciting.

KaLok Man
Nan Zhang
Dawei Liu
Dejun Xie
DATICS-MUE'14 IPC Chairs

DATICS-2014 Organization

Program Chairs

KaLok Man	Xi'an Jiaotong-Liverpool University, China
Nan Zhang	Xi'an Jiaotong-Liverpool University, China
Dawei Liu	Xi'an Jiaotong-Liverpool University, China

Organizing Chairs

Michele Mercaldi	EnvEve, Switzerland
Chi Un Lei	University of Hong Kong
Tomas Krilavicius	Baltic Institute of Advanced Technologies and Vytautas Magnus University, Lithuania
DejunXie	South University of Science and Technology, China

Program Committee

Vladimir Hahanov	Kharkov National University of Radio Electronics, Ukraine
Paolo Prinetto	Politecnico di Torino, Italy
Massimo Poncino	Politecnico di Torino, Italy
Alberto Macii	Politecnico di Torino, Italy
Joongho Choi	University of Seoul, South Korea
Wei Li	Fudan University, China
Michel Schellekens	University College Cork, Ireland
Emanuel Popovici	University College Cork, Ireland
Jong-Kug Seon	System LSI Lab., LS Industrial Systems R&D Center, South Korea
Umberto Rossi	STMicroelectronics, Italy
Franco Fummi	University of Verona, Italy
Graziano Pravadelli	University of Verona, Italy

Yui Fai Lam	Hong Kong University of Science and Technology, Hong Kong
Ajay Patel	Intelligent Support Ltd, United Kingdom
Jinfeng Huang	Philips & LiteOn Digital Solutions Netherlands, The Netherlands
Thierry Vallee	Georgia Southern University, Statesboro, Georgia, USA
Monica Donno	Minteos, Italy
Jun-Dong Cho	Sung Kyun Kwan University, South Korea
AHM ZahirulAlam	International Islamic University Malaysia, Malaysia
Gregory Provan	University College Cork, Ireland
Miroslav N. Velev	Aries Design Automation, USA
M. Nasir Uddin	Lakehead University, Canada
Dragan Bosnacki	Eindhoven University of Technology, The Netherlands
Milan Pastrnak	Siemens IT Solutions and Services, Slovakia
John Herbert	University College Cork, Ireland
Zhe-Ming Lu	Sun Yat-Sen University, China
Jeng-Shyang Pan	National Kaohsiung University of Applied Sciences, Taiwan
Chin-Chen Chang	Feng Chia University, Taiwan
Mong-Fong Horng	Shu-Te University, Taiwan
Liang Chen	University of Northern British Columbia, Canada
Chee-Peng Lim	University of Science Malaysia, Malaysia
Salah Merniz	Mentouri University, Constantine, Algeria
Oscar Valero	University of Balearic Islands, Spain
Yang Yi	Sun Yat-Sen University, China
Damien Woods	University of Seville, Spain
Franck Vedrine	CEA LIST, France
Bruno Monsuez	ENSTA, France
Kang Yen	Florida International University, USA
Takenobu Matsuura	Tokai University, Japan
R. Timothy Edwards	MultiGiG, Inc., USA
Olga Tveretina	Karlsruhe University, Germany
Maria Helena Fino	Universidade Nova De Lisboa, Portugal
Adrian Patrick ORiordan	University College Cork, Ireland
Grzegorz Labiak	University of Zielona Gora, Poland
Jian Chang	Texas Instruments, Inc, USA
Yeh-Ching Chung	National Tsing-Hua University, Taiwan
Anna Derezinska	Warsaw University of Technology, Poland
Kyoung-Rok Cho	Chungbuk National University, South Korea
Yuanyuan Zeng	Wuhan university, China
D.P. Vasudevan	University College Cork, Ireland

Arkadiusz Bukowiec	University of Zielona Gora, Poland
Maziar Goudarzi	Sharif University of Technology, Iran
Jin Song Dong	National University of Singapore, Singapore
Dhamin Al-Khalili	Royal Military College of Canada, Canada
Zainalabedin Navabi	University of Tehran, Iran
Lyudmila Zinchenko	Bauman Moscow State Technical University, Russia
Muhammad Almas Anjum	National University of Sciences and Technology (NUST), Pakistan
Deepak LaxmiNarasimha	University of Malaya, Malaysia
Danny Hughes	Katholieke Universiteit Leuven, Belgium
Jun Wang	Fujitsu Laboratories of America, Inc., USA
A.P. Sathish Kumar	PSG Institute of Advanced Studies, India
N. Jaisankar	VIT University, India
Atif Mansoor	National University of Sciences and Technology (NUST), Pakistan
Steven Hollands	Synopsys, Ireland
Siamak Mohammadi	University of Tehran, Iran
Felipe Klein	State University of Campinas (UNICAMP), Brazil
Enggee Lim	Xi'an Jiaotong-Liverpool University, China
Kevin Lee	Murdoch University, Australia
Prabhat Mahanti	University of New Brunswick, Saint John, Canada
Kaiyu Wan	Xi'an Jiaotong-Liverpool University, China
Tammam Tillo	Xi'an Jiaotong-Liverpool University, China
Yanyan Wu	Xi'an Jiaotong-Liverpool University, China
Wen Chang Huang	Kun Shan University, Taiwan
Masahiro Sasaki	The University of Tokyo, Japan
Shishir K. Shandilya	NRI Institute of Information Science & Technology, India
J.P.M. Voeten	Eindhoven University of Technology, The Netherlands
Wichian Sittiprapaporn	Mahasarakham University, Thailand
Aseem Gupta	Freescale Semiconductor Inc., Austin, TX, USA
Kevin Marquet	Verimag Laboratory, France
Matthieu Moy	Verimag Laboratory, France
Ramy Iskander	LIP6 Laboratory, France
Suryaprasad Jayadevappa	PES School of Engineering, India
Shanmugasundaram Hariharan	PavendarBharathidasan College of Engineering and Technology, India
Chung-Ho Chen	National Cheng-Kung University, Taiwan
Kyung Ki Kim	Daegu University, Korea
Shiho Kim	Chungbuk National University, Korea
Hi Seok Kim	Cheongju University, Korea

Brian Logan	University of Nottingham, UK
Asoke Nath	St. Xavier's College (Autonomous), India
Tharwon Arunuphaptrairong	Chulalongkorn University, Thailand
Shin-YaTakahasi	Fukuoka University, Japan
Cheng C. Liu	University of Wisconsin at Stout, USA
Farhan Siddiqui	Walden University, Minneapolis, USA
Katsumi Wasaki	Shinshu University, Japan
Pankaj Gupta	Microsoft Corporation, USA
Taikyeong Jeong	Myongji University, South Korea
Masoud Daneshtalab	University of Turku, Finland
Amit Chaudhry	Technology Panjab University, India
Bharat Bhushan Agarwal	I.F.T.M., University, India
Abhilash Goyal	Oracle (SunMicrosystems), USA
Boguslaw Cyganek	AGH University of Science and Technology, Poland
Yeo Kiat Seng	Nanyang Technological University, Singapore
Youngmin Kim	UNIST Academy-Industry Research Corporation, South Korea
Tom English	Xlinx, Ireland
Nicolas Vallee	RATP, France
Mou Ling Dennis Wong	Xi'an Jiatong Liverpool University, China
Rajeev Narayanan	Cadence Design Systems, Austin, TX, USA
Xuan Guan	Freescale Semiconductor, Austin, TX, USA
Pradip Kumar Sadhu	Indian School of Mines, India
Fei Qiao	Tsinghua University, China
Chao Lu	Purdue University, USA
Ding-Yuan Cheng	National Chiao Tung University, Taiwan
Pradeep Sharma	IEC College of Engineering & Technology, Greater Noida, GB Nagar UP, India
Ausra Vidugiriene	Vytautas Magnus University, Lithuania
Sheung-Hung Poon	National Tsing Hua University, Taiwan
Lixin Cheng	Suzhou Institute of Nano-Tech and Nano-Bionics (SINANO), Chinese Academy of Sciences, China
Yue Yang	Suzhou Institute of Nano-Tech and Nano-Bionics (SINANO), Chinese Academy of Sciences, China
Yo-Sub Han	Yonsei University, South Korea
Chien-Chang Chen	Tamkang University, Taiwan
Hui-huang Hsu	Tamkang University, Taiwan
Hwann-Tzong Chen	National Tsing Hua University, Taiwan

Message from FSTA-14 Workshop Chair

Welcome to the 2014 International Workshop on Future Science Technologies and Applications (FSTA 2014).

The Internet, the cellular and the wireless systems are converging now, thus giving birth to the future convergence Internet. The International Workshop on FSTA 2012, 2013 brings together scientists, engineers, computer users, and students to exchange and share their experiences, new ideas, and research results about all aspects (theory, applications and tools) of computer and information science, and discuss the practical challenges encountered and the solutions adopted. The workshop on 'future science technologies and applications' aims to serve as an international forum for researchers and practitioners willing to present their early research results and share experiences in the field of future science and computer.

FSTA-2014 contains high quality research papers submitted by researchers from all over the world. Each submitted paper was peer-reviewed by reviewers who are experts in the subject area of the paper. Based on the review results, the Program Committee accepted 5 papers.

For organizing an International workshop, the support and help of many people is needed. We would like to thank all authors for their work and presentation, all members of the program committee and reviewers for their cooperation and time spent in the reviewing process. Particularly, we thank the founding steering chair of MUE 2014, James J. (Jong Hyuk) Park, Workshop chair MUE 2014, and Program chair of MUE 2014. Finally, Special thanks are extended to the staffs from FSTA 2014, who contributed so much to the success of the conference.

FSTA-2014 Workshop Program Chairs
Namje Park

FSTA-2014 Organization

Steering Chair

James J. Park	SeoulTech, Korea

General Chairs

Jianhua Ma	Hosei University, Japan
Shu-Ching Chen	Florida International University, USA
Timothy K. Shih	National Central University, Taiwan
Doo-soon Park	SoonChunHyang University, Korea
Qingguo Zhou	Lanzhou University, China

Program Chairs

Joon-Min Gil	Catholic University of Daegu, Korea
Neil Y. Yen	The University of Aizu, Japan
Shingchern You	National Taipei Universiy of Technology, Taiwan

Publicity Chairs

Kehua Guo	CSU, China
Mohamed Gaber	University of Portsmouth, UK
Jason C. Hung	Oversea Chinese University, Taiwan
Ryan Leong Hou U	Universidade De Macau, China
Chengjiu Yin	Kyushu University, Japan
Jian-Lian Chen	Aletheia University, Taiwan
Junbo Wang	University of Aizu, Japan
Bong-Hwa Hong	Kyung Hee Cyber University, Korea
Nan-Chen Hsieh	National Taipei University and Health Sciences, Taiwan
Deok-Gyu Lee	ETRI, Korea
Young-Ae Jeong	Sunmoon University, Korea

Program Committee

Qiang Zhu	The University of Michigan, USA
Shao-Shin Hung	WuFeng University, Taiwan
Tat-Chee Wan	UniversitiSains Malaysia
Jun Xiao	East China Normal University, China
Young soo Kim	Electronics and Telecommunications Research Institute, Korea
Oliver Amft	Eindhoven University of Technology, Potentiaal
Changjing Shang	Aberystwyth University, UK
Wanqing Tu	Glyndwr University, UK
Kok-Seng Wong	Soongsil University, Korea
Narayanan Kulathuramaiyer	Universiti Malaysia Sarawak, Malaysia
Nik Bessis	University of Derby, UK
Qinghe Du	Xi'an Jiaotong University, China
Kyoil Cheong	Electronics and Telecommunications Research Institute, Korea
Rachid Anane	Coventry University, UK
Mehul Bhatt	University of Bremen, Germany
Zhou Su	WasedaUniv., Japan
Chee-Peng Lim	Deakin University, Australia
Rami Yared	Japan Advanced Institute of Science and Technology, Japan
M. Elena Renda	IIT-CNR, Italy
Dooho Choi	Electronics and Telecommunications Research Institute, Korea
Youngsook Lee	Howon University, Korea
Soohyun Oh	Hoseo University, Korea
Taesung Lee	KISA, Korea

Contents

Multimedia Modeling and Processing

Ubiquitous and Pervasive Computing

Artificial Intelligence Applications for E-services (AIAE)

Design, Analysis and Tools for Integrated Circuits and Systems (DATICS)

Future Security Technologies and Applications (FSTA)

3D Mapping of Garment Patches Based on Human Body Section Loop Data

Lin Xiaodan, He Bing, and Zhang Xiulin

State Key Lab of Virtual Reality Technology and Systems,
Beijing, China
amylindan@gmail.com

Abstract. Virtual-try-on technology has been gaining popularity for its commercial potential. Many researches implement 3D garments mapping by sewing virtual garment patterns together around a mannequin. However, many systems need user interaction. In this paper, we propose a 3D mapping method of garment patches based on human body section loop data. The inputs include a virtual mannequin and the front and back patterns of a garment. The virtual-try-on process needs no user interference and can satisfy the real-time requirement after the pretreatment of human body section loop data.

Keywords: virtual-try-on, section loop data, garment locating loop, 3D mapping.

1 Introduction

The Internet becomes an attractive channel for the sale of garment products nowadays. With the development of virtual garment design and simulation techniques, virtual-try-on visualization has emerged as a compelling tool for online garment sales.

Many virtual-try-on system prototypes have been proposed and different simulation techniques have been applied so far. Meng et al.[1] and Wacker et al.[2] fulfilled 2D garment mapping to 3D surface by sewing virtual garment patterns together around a mannequin. Divivier et al.[3] proposed a set of methods for geometric pre-positioning, modeling and simulation. Ding et al.[4] put forward a co-evolutionary immune algorithm and use it to solve the large scale garment matching problem.

Generally, the typical virtual-try-on systems involve five processes including creation of human avatar (namely a 3D virtual counterpart of the user), 2D garment pattern design, pre-positioning of garment patterns, a virtual sewing process and physical simulations. Among the five processes, pre-positioning of garment patterns is a critical step. The result of pre-positioning decides the virtual-try-on effect to a great extent. However, there is no best method for automatic pre-positioning, and most systems complement pre-positioning by user interaction.

In this paper, we propose a 3D mapping method of garment patches based on human body section loop data. The inputs include a virtual mannequin and the front and back patterns of a garment. The virtual-try-on process needs no user interference and can satisfy the real-time requirement after the pretreatment of human body section loop data. Researches have been done for 3D body measurement and body scanning techniques concerning creation of avatar which is beyond the scope of this paper[5,6].

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_1, © Springer-Verlag Berlin Heidelberg 2014

2 Pretreatment of the Human Body Section Loop Data

A virtual 3D mannequin is generally expressed by triangle meshes, which seems redundant in the human body information expression and increases calculation workload of garment 3D mapping. In this paper, the mannequin data is organized based on the human body section loops. The loops are produced from the mannequin expressed by triangle meshes.

2.1 Generation of the Human Body Section Loop Data

We divide the mannequin into several parts including trunk, left arm, left forearm, right arm, right forearm, left thigh, left shank, right thigh and right shank. For each part, several cross sections are used to cut the mannequin and a slice is obtained when a cross section intersects with the mannequin part. Each slice is formed by connecting the intersection point set and these slices form the human body section loop data.

2.2 Demarcation of Vertical Marker Lines of the Mannequin

Four marker lines are demarcated on the mannequin. Two of them are on the front side and the other two are on the back side. The four lines all start from neck, and the front two pass through each side of breast while the back two pass through scapulae respectively, finally all of them extend downward to shanks. As marker lines of different people vary, they are irregular. Thus we demarcate them manually.

3 The Garment Locating Loops

In real life, there are gaps between a garment and a mannequin. For instance, as the chest circumference of a woman is greater than waistline, the waist part of a garment is always pending. To get pending effect, the human body section loop data is adjusted and the adjusted result is defined as garment locating loops.

3.1 The Adjusting Algorithm

The adjusting algorithm is described as follows, and the torso is a major part to be adjusted. Adjust the section loop data successively along the torso axis, starting from the first loop at neck. The Z axis is the torso axis in Fig. 1(a).

Definition 1(Current Loop, L_{cur}).
The current loop refers to the loop needed to be adjusted in this algorithm. In Fig. 1(a), the lower loop is the current loop.

Definition 2(Contour Loop, L_{con}).
Project the upper loop onto the plane of lower loop along the negative direction of Z axis, then contour loop is defined as the loop encircles the lower loop and the projection of the upper loop tightly. There are three circumstances as shown in Fig. 1(a). When one loop contains the other loop, the outer loop is taken as the contour loop. When the two loops intersect, the encircling loop is taken.

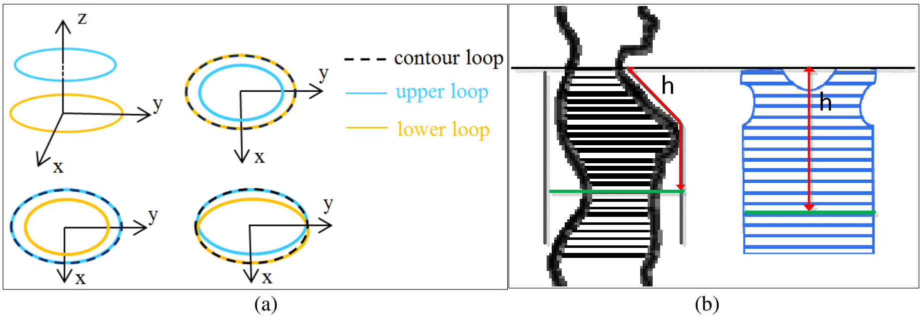


Fig. 1. (a) Contour loop (b) The measuring method of the distances

Definition 3(Corresponding Weft Length, CWL).
The corresponding weft length refers to the length of the garment weft whose distance from shoulder seam is the same as the distance between the current loop and the first loop at neck. The measurement of the distances is shown in Fig. 1(b). The distance between the current loop and the first loop at neck is measured along the vertical marker lines of the mannequin.

Adjusting Algorithm:
Define: **C**(L) is the circumference of loop L.
If CWL < C(L_{cur})
  Then L_{cur} needs no adjustment
Else if CWL > C(L_{cur}) && CWL < C(L_{con})
  Then extend L_{cur} until its circumference reaches CWL
Else if CWL > C(L_{con})
  Then extend L_{cur} until its circumference reaches C(L_{con})

3.2 Optimization of the Garment Locating Loops

The correctness of the garment locating loop data will directly affect the effect of the 3D garment simulation. Therefore, the garment locating loops should be optimized before 2D-3D patch mapping.

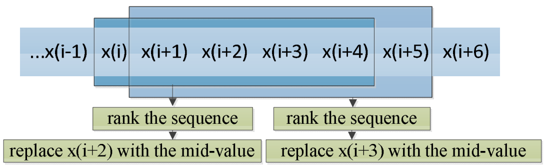


Fig. 2. One-dimensional median filter method

The garment locating loops optimization can be seen as a process of re-sampling and smoothing of curve in the plane. In this paper, one-dimensional median filter method is adopted to smooth the curve, and the basic principle of the method is to replace the value at the central point in digital sequence with mid-value in the point domain as shown in Fig. 2. The specific steps are described as follows:

(1) Each garment locating loop can be seen as a point set on the plane, or a digital sequence. A median filter window with a length of 2n+1 is set up for these point sets and moved along the digital sequence with a step length of 1 each time. In our experiment, n is taken as 2;
(2) After moving the window, sequences in the window should be ranked;
(3) Calculate the mid-value of the sequence in the window and replace the value at the central point with mid-value;
(4) Repeat until the end of the point sequence.

4 2D-3D Mapping of Garment Patch

4.1 The Garment Locating Loops Unfolding

Unfolding of the garment locating loops is designed to facilitate 2D-3D mapping. The principle of the garment locating loops unfolding algorithm is that the loop data are considered as a 3D point set, and one point has relative positional relation with the other point, as long as the loop data after unfolding can keep this relation to the maximum extent, the unfolding algorithm is regarded as rational.

Constraint conditions to be considered during the unfolding process include spatial distance and curvature constraint. Azariadis[7] et al. put forward a curvature unfolding algorithm, in which, distance constraint equals to constraint of the sum of the distances between one point and other vertices of triangle with the point as the vertex, while the curvature constraint equals to the constraint of the included angle formed by edges of triangle.

In the algorithm proposed by Azariadis, when the distances constraint is considered, the way of curved surface triangularization influences the number of affected vertices. If the triangular mesh is irregular, the number of vertices to be taken into consideration will be less, thus losing some curved surface information and leading to inaccurate curved surface unfolding. Fig. 3 shows the relationship between the way of curved surface triangularization and the number of affected vertices. In this paper, the original algorithm is improved, distance constraint is no more considered based on the adjacent triangle after the curved surface triangularization. No matter how the curved surface is triangularized, the spatial distance constraint should ensure that eight points near each point (boundary excluded) are taken into consideration. In our experiment, conjugate gradient method is adopted to calculate the minimum value.

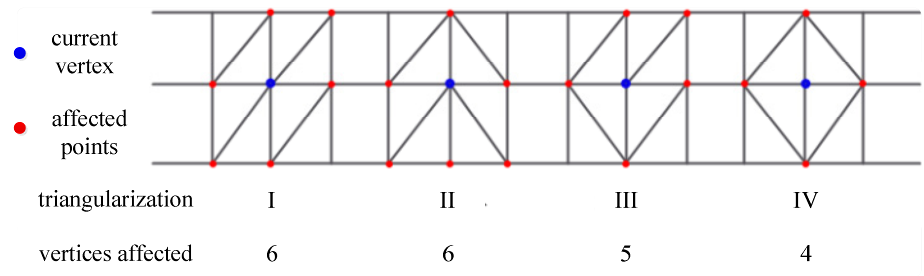


Fig. 3. Relation between triangularization method and number of affected vertices

4.2 2D-3D Garment Patch Mapping

The unfolded garment locating loops and the 2D garment patch are on the same plane and the unfolded garment locating loops contain both planar and 3D coordinates. Therefore, we can implement the 2D to 3D garment patch mapping as follows:

(1)The garment patch is subdivided into mesh grids using regular grid method as shown in Fig. 4.

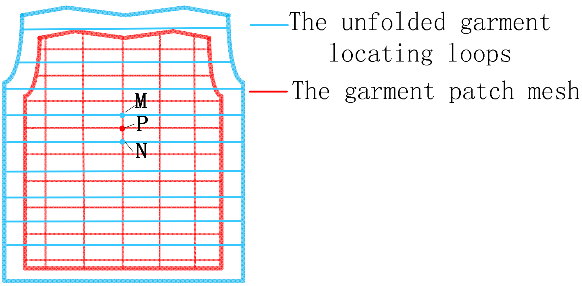


Fig. 4. 2D-3D garment patch mapping

(2)We let the unfolded garment locating loops overlap the 2D garment patch mesh by putting them in the same coordinates.

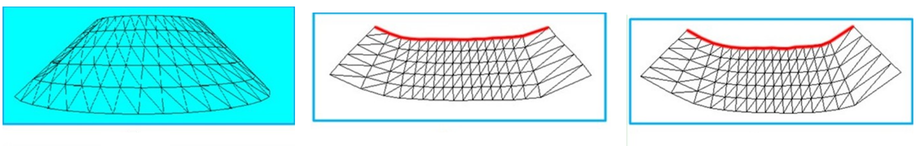
(3) By matching the 2D coordinates of the unfolded garment locating loops with that of 2D garment patch mesh, 3D coordinates of the garment patch mesh vertices can be calculated by the 3D coordinates of the unfolded garment locating loops. As shown in Fig , the 3D coordinate of vertex P on the garment patch mesh can be interpolated by the 3D coordinates of vertices M and N on the unfolded garment locating loops.

Using the method described above, 2D-3D garment patch mapping is completed and a 3D garment wore on a mannequin is acquired.

5 Experiment Results

5.1 Experiment I: Comparison between the Unfolding of Curved Surface Adopting the Azariadis's Algorithm and the Improved Algorithm

Fig. 5 shows the comparison result. The spatial curved surface to be unfolded in this experiment contains 108 vertices and 170 triangles; curvature weight in the unfolding algorithm is 0.9, and distance weight is 0.1. According to the figures, after improving



(a) (b) (c)

Fig. 5. (a) Spatial curved surface to be unfolded; (b) Effect of curved surface unfolded adopting the Azariadis's algorithm; (c) Effect of curved surface unfolded using the improved algorithm

the unfolding algorithm, boundary lines of the unfolded curved surface are more curved, with less information lost and accurate data obtained.

5.2 Experiment II: 3D Garment Simulation Result

Our method has good simulation effect. The overall simulation effects of a shirt and a trouser are shown in Fig. 6, and both of the simulation processes take less than 5 seconds.

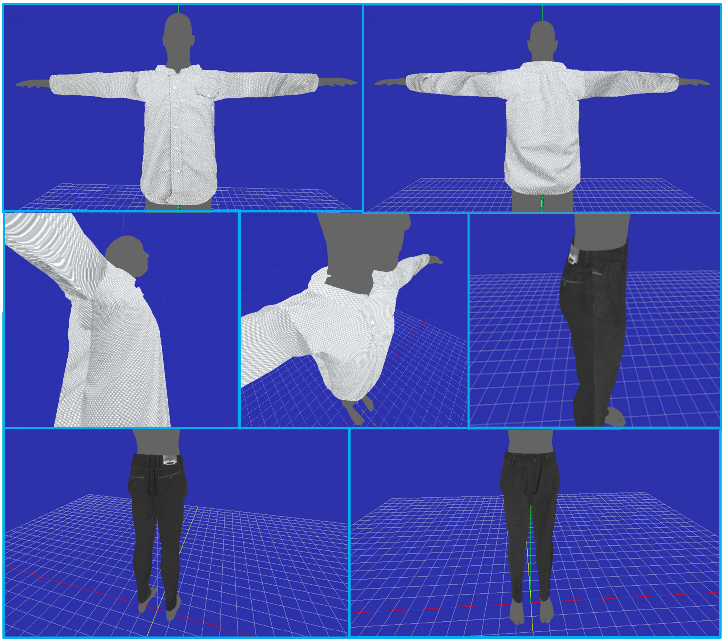


Fig. 6. Coat/trouser simulation effect

6 Summary

A garment patch 3D mapping method based on the human body section loop data is presented in the paper. The human body model is expressed by the section loop data. First, a virtual mannequin expressed by triangle meshes is cut into section loops, and vertical marker lines of the mannequin are demarcated on the resulted section loop data. Then, garment locating loops are generated from section loops. Finally, 2D-3D garment patch mapping is completed by calculating 3D coordinates of the garment patch mesh vertices from the information on unfolded garment locating loops. Experiments show that the section loop data have obvious advantages in extracting human body characteristic information. The method proposed is totally geometry-based. To improve visual effects, physical details would be added in our future work.

Acknowledgments. This work was supported by grant No. 61272346 from NSFC (National Natural Science Foundation of China).

References

1. Meng, Y., Mok, P.Y., Jin, X.: Interactive virtual try-on clothing design systems. J. Computer-Aided Design 42(4), 310–321 (2010)
2. Wacker, M., Keckeisen, M., Kimmerle, S., et al.: Simulation and Visualization of Virtual Textiles for Virtual Try-On. J. Special Issue of Research Journal of Textile and Apparel: Virtual Clothing Technology and Applications 9(1), 37–47 (2005)
3. Divivier, A., Trieb, R., Ebert, A., et al.: Virtual try-on topics in realistic, individualized dressing in virtual reality. J. (2004)
4. Ding, Y.S., Hu, Z.H., Zhang, W.B.: Multi-criteria decision making approach based on immune co-evolutionary algorithm with application to garment matching problem. J. Expert Systems with Applications 38(8), 10377–10383 (2011)
5. Dai, L.: The 3D digital technology of fashion design. In: 2011 International Symposium on Computer Science and Society (ISCCS), pp. 178–180. IEEE (2011)
6. Yu, X., Wang, J.: The Matching Principle of Bodily Form and Garment Size Series and Its Application in Garment Electronic Commerce. J. Journal of Donghua University, Natural Science, 1–9 (2003)
7. Azariadis, P.N., Aspragathos, N.A.: Geodesic curvature preservation in surface flattening through constrained global optimization. J. Computer-Aided Design 33(8), 581–591 (2001)

A Robust Hand Tracking Approach Based on Modified Tracking-Learning-Detection Algorithm

Haibin Shi[1], Zhanjian Lin[1], Weiwei Tang[1], Bruce Liao[2], Jolly Wang[2], and Lingxiang Zheng[1,*]

[1] School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
{shihaibin,lxzheng}@xmu.edu.cn
[2] Intel (China) Co., Ltd., Shanghai 200241, China
{bruce.liao,jolly.wang}@intel.com

Abstract. Hand tracking is an essential step for dynamic gesture recognition which catches a lot of attention in the field of gesture interaction. In this paper, we present a robust hand tracking approach for unconstrained videos based on modified Tracking-Learning-Detection (TLD) algorithm, named BP-TLD. By introducing a skin color feature to the model, we make the algorithm more suitable for hand tracking. The experimental results show that BP-TLD has a better performance compared with other tracking algorithms such as TLD, MSEPF and Handvu. It indicates that our approach can meet the requirements of robustness and real-time better for the frontal-view vision-based human computer interactions.

Keywords: Back Projection, Tracking-Learning-Detection, Hand Tracking.

1 Introduction

Hand tracking is an essential step for dynamic hand gesture recognition. Considering the natural gesture tracking and the current common hardware configuration, monocular computer vision of gesture tracking catches a lot of attention in the field of gesture interaction.

Several approaches have been used in hand tracking, such as Selma Belgacem et al. [1] proposed an idea of using Optical-Flow embedded Particle Filter in sign language scenes. It improved the tracking precision of gestures. However, the main characteristic of Particle Filter is color histogram which is restricted by the accuracy of skin region extraction and skin-color interference. Kolsch [2] proposed an idea that we can track flock of features to locate the hand through a head-mounted camera. This method assumed that the hand region dimension was fixed and not many other objects had the similar color with the skin in the background. Stenger et al. [3] proposed a gesture tracking process in accordance with the speed condition of the background and speed of movement of the hand. Adaptive selection gesture tracking uses one or several features such as color, movement, etc. This approach can choose suitable algorithms on the basis of different circumstance to guarantee the accuracy and robustness.

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_2, © Springer-Verlag Berlin Heidelberg 2014

The algorithms above rely on skin-color information seriously. However, since the skin-color is an extraordinary unstable factor, it is difficult for the system to guarantee the accuracy of skin-color area under interferential circumstance. Zdenek proposed the Tracking-Learning-Detection (TLD) algorithm [4-6] during his PhD in University of Surrey, which is an excellent algorithm that aims at long-term tracking of unknown objects. But some problems remain to be solved when apply it to tracking gesture, among which the relatively slow speed of the system due to the complex algorithm and great calculation cost is a priority. In this paper, we try to improve the accuracy of real-time hand tracking based on a modified Tracking-Learning-Detection (TLD) algorithm.

This paper is organized as follows: Section two gives an introduction to the BP-TLD algorithm. Section three presents the experimental results for different video sequences. Conclusions are made in Section four.

2 BP-TLD Hand Tracking

The framework of the BP-TLD hand tracking algorithm is shown in figure 1. Our purpose is to improve the tracking speed by reducing the calculation cost of TLD algorithm. We evaluate the skin color information of the image by calculating the Back Projection image when the system tracks the gesture area successfully. Then we try to build a skin color classifier by making use of the generated Back Projection image to filter the detection windows in the case of the little interference of the environment before using the TLD algorithm. The detailed process of the framework will be discussed in the subsequent part of the section.

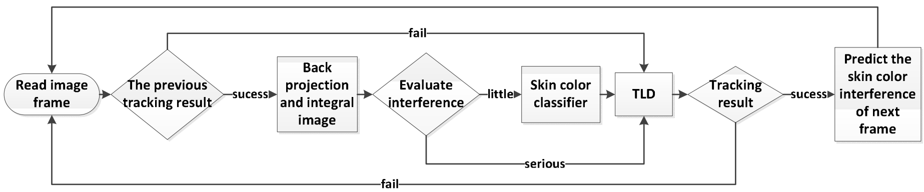


Fig. 1. The framework of the BP-TLD

2.1 Tracking-Learning-Detection

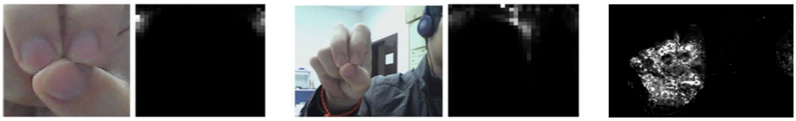
TLD is a robust object tracking method based on selected region matching. Compared to the traditional track algorithm, TLD decomposes the long-term tracking task into tracking, learning and detection explicitly. Based on Median-Flow tracker [7] extended with failure detection, the tracker estimates the motion of object between consecutive frames. The detector treats every frame as individual and localizes all the appearances that have been observed so far and learned in the past. The learning module initializes the object detector in the first frame and estimates the errors of detector through a novel learning method (P-N learning) [8]. P-N learning defines two types of “experts”: P-expert identifies only false negatives and N-expert identifies only false positives.

As an important part of TLD, object detector scans the input image by a scanning-window and each patch determines the presence or absence of the object. According to an initial bounding box, the scanning-window generates all possible scales and shifts of the box. A large number of bounding boxes are needed to be evaluated. In this case, an efficient cascaded classifier [9] is structured into three stages: patch variance, ensemble classifier and nearest neighbor. Each stage either refuses the patch in question or passes it to the next stage. Templates are used to estimate the reliability of the detection more effectively. Thus, it is clear that the detector of TLD accounts for a majority of total calculation.

2.2 Back Projection

Back Projection is a method to record how well the pixels of a given image fit the distribution of pixels in a histogram model. In other words, Back Projection is calculating the histogram model of a feature and then using the model to find this feature in another image. The Back Projection can be described as follows:

Suppose we have a skin histogram (Hue-Saturation) as the image below (Fig.1 (a)). The image (Fig.1 (a)-right) is the model histogram which represents the skin tonality of the sample. Then you can capture the histogram of the skin area by mask operation:



(a) Feature gesture and its Hue-Saturation histogram (b) Test Image and its Hue-Saturation histogram (c) skin probability image

Fig. 2. Feature gesture and Test Image

The skin histogram is a three-dimensional data graph, in which x-y represents the two- dimensional H-S color space and the value stored in (x,y) represents the probability of the color ever appeared. It can be calculated by:

$$v(h,s) = \frac{num(h,s)}{sum} \tag{1}$$

v(h, s) is the value of the model histogram in position (h, s) ,as we can see, v(h,s) equals the number of pixels with hue and saturation equal h and s in the image of feature gesture divides the number of pixels in the image of feature gesture. A crucial step in the analysis process is to use the model histogram to detect skin areas in the Test Image. The transform equation is:

$$f(x_i, y_j) = v(h,s) \tag{2}$$

Where (h, s) are the values of the hue and saturation of pixel (x_i,y_j) in the Test Image, $f(x_i,y_j)$ is the value of pixel (x_i,y_j) in the new image which we call the Back Projection image (Fig.1 (c)) for the Test Image. Statistically speaking, the values represent the probability that a pixel in Test Image belongs to skin area. In Back Projection image, the brighter areas are more probable to be skin areas.

2.3 BP-TLD

Based on the Back Projection algorithm, we introduce the skin-color classifier for TLD algorithm. Since the color information is impressionable, the skin-color classifier should be designed based on interference prediction and frame coherence.

Here we describe the design procedures of BP-TLD algorithm in details:

①Read image frame: First of all, the system obtains a frame from camera and determines whether to use the skin-color classifier by judging the tracing result of the previous frame. If the previous frame tracks successfully, the system will judge interference situation. Otherwise, it will skip step ④.

②Calculate Back Projection and integral image: Obtain skin probability map by equation (1) for the classifier.

③Evaluate the mutation interference which is mainly caused by the camera and light condition:

$$ratio = \frac{\sum_i \sum_j f_c(x_i, y_j)}{\sum_i \sum_j f_p(x_i, y_j)} \tag{3}$$

Where $f_c(x_i,y_j)$ is the value of pixel (x_i,y_j) in the Back Projection image of the current frame and $f_p(x_i,y_j)$ is that of the previous frame. If the ratio is greater than the empirical value, it is assumed that this frame is interfered severely and we jump to step ⑤ directly.

④Skin-color classifier: We calculate the rate of the skin color area in the alternative window:

$$ratio = \frac{\sum_i \sum_j f_a(x_i, y_j)}{M_a * N_a} \tag{4}$$

Where $f_a(x_i,y_j)$ is the value of pixel (x_i,y_j) in the Back Projection image of the alternative window and Ma*Na is the size of this window. If this ratio is less than the threshold from the previous frame, we discard this window, otherwise retain it.

⑤Track by TLD: We input the retained window to the cascade classifier of TLD detector. After TLD tracking, the system can adjust the flag of classifier according to the tracking result.

⑥Calculate the threshold of skin-color: The color threshold is the ratio multiplied by a coefficient. And the ratio is calculated by the formula (4).The coefficient is also an empirical value of 0.66 through the observation of our test video.

⑦Predict the skin-color interference of next frame:

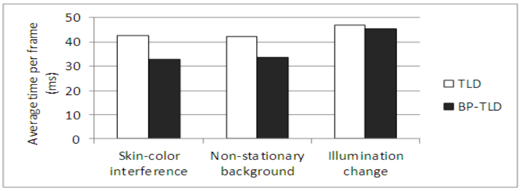
$$ratio = \frac{\sum_i \sum_j f_t(x_i, y_j)}{\sum_i \sum_j f(x_i, y_j)} \tag{5}$$

Where $f_t(x_i,y_j)$ is the value of pixel (x_i,y_j) in the Back Projection image of the target window and $f(x_i,y_j)$ is that of the whole frame. If the value is less than the empirical value (set as 0.25) that means the system is seriously disturbed by skin color object and should skip the skin-color classifier in the next frame, otherwise the result is acceptable. Read the next frame and analyze it.

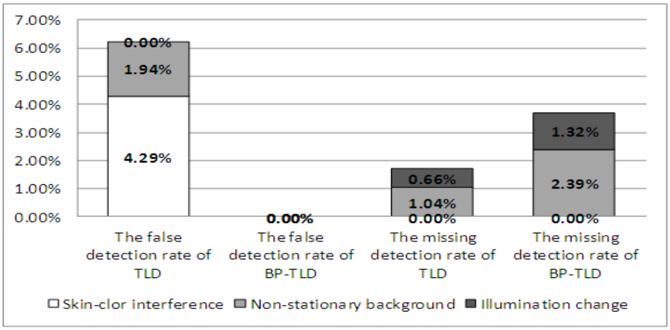
3 Experimental Results and Analysis

In order to compare the efficiency and accuracy between BP-TLD and TLD, three video sequences are offered for the test. The first is the interference of the skin-color object. We use a chunk of wood as skin-color distractions. The second is the interference of non-stationary background where volunteer moves around freely. The last one is the interference of strong backlight environment where the white wall reflects the sunlight at noon. These video sequences are recorded between 685 and 906 frames. The camera captures 320*240 video at 30 frames a second. We run our algorithm on a common PC, which has an Intel i5 650 CPU and 2G memory.

Three video sequences, catalogued by skin-color interference, non-stationary background and illumination change, are recorded to test the efficiency of BP-TLD. Each video has about 700-900 frames.



(a) Comparison of tracking time



(b) Comparison of false detection rate and missing detection rate

Fig. 3. Performance Comparison of BP-TLD and TLD

Shown as Fig.3(a), the successful tracking time of BP-TLD reduces ten milliseconds per frame, which shows that the proposed algorithm has a 20% performance improvement compared to the original TLD. And Fig.3(b) suggests that BP-TLD increases not only the tracking speed but also the overall tracking success rate. Especially in all kinds of interference, BP-TLD maintains a high level of the zero error detection rate, although missing detection rate increases. This is because we use the color information. It improves the accuracy of tracking, makes online learning samples with higher quality and reduces the probability of false detection. On the other hand, the color information is not stable. When interfered, it can lead to missing tracking. In a word, the results show that the BP-TLD algorithm which has optimized the detector significantly improves the tracking speed and is more applicable.

Table 1. Test video sequences

ID	The speed of hand motion	Background complexity	Total frames
1	Slow	few skin color object(single people)	927
2	Slow	complicated(two people, wooden furniture)	954
3	Fast	few skin color object(single people)	897
4	Fast	complicated(two people, wooden furniture)	790

To test our algorithm, we compare our tracking algorithm with MSEPF algorithm [10] and Handvu algorithm [11]. MSEPF is based on mean shift embedded particle filter and Handvu is based on the non-weighted flocks of KLT features. Four videos with the size of 640*480 are recorded for experiment according to hand motion speed and background complexity. The result is shown in the Fig.4.

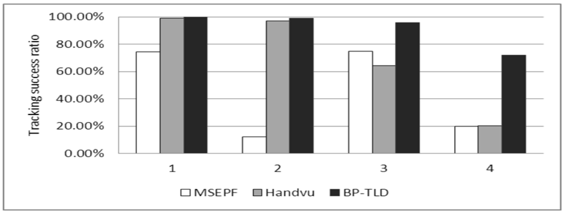


Fig. 4. Comparing with other methods

The result indicates that our algorithm is better than MSEPF and Handvu. When the number of skin color objects increases in the scene, the MSEPF cannot track the hand very well. For MSEPF uses H-S histogram to find hands, it cannot deal with complicated background. Under the condition of slow hand motion (video 1, 2), Handvu and our algorithm have a better results. But if we improve the speed of hand motion, our approach outperforms Handvu. The result reflects that the illumination variations which caused by fast motion can affect the optical flow tracking of Handvu, and it is hard to find the hand again. On the other hand, the lack of adaptive mechanism in the framework of MSEPF and Handvu also make them easily fail to track.

4 Conclusion

In this paper, the BP-TLD algorithm witch is a robust real-time hand tracking approach based on modified Tracking-Learning- Detection algorithm is introduced. To reduce the interference of some complex background, especially when there are lots of skin color objects, we define a particular gesture tracked by BP-TLD. On the other hand, in order to improve the tracking rate, back projection algorithm is used for the detector of

BP-TLD, thus the fast moving hand with particular gesture can be robustly tracked even in cluttered background. Experimental results show that BP-TLD algorithm has a better performance than the original TLD algorithm. And from the comparison with other hand tracking algorithms, we can find that the successful tracking rate of our BP-TLD is several times better than MSEPF and Handvu. Compared with binocular stereo vision and Kinect, our algorithm is economic, simple and easy to popularize. In the future, it can be developed into multi-target gesture tracking with higher practicability.

References

1. Belgacem, S., Chatelain, C., Ben-Hamadou, A., Paquet, T.: Hand Tracking Using Optical-Flow Embedded Particle Filter in Sign Language Scenes. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 288–295. Springer, Heidelberg (2012)
2. Kolsch, M., Turk, M.: Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration. In: Computer Vision and Pattern Recognition Workshop, Washington, pp. 158–158 (2004)
3. Stenger, B., Woodley, T., Cipolla, R.: A vision-based remote control. In: Cipolla, R., Battiato, S., Farinella, G.M. (eds.) Computer Vision. SCI, vol. 285, pp. 233–262. Springer, Heidelberg (2010)
4. Kalal, Z., Matas, J., Mikolajczyk, K.: Weighted sampling for large-scale boosting (2008)
5. Kalal, Z., Mikolajczyk, K., Matas, J.: Face-tld: Tracking-learning-detection applied to faces. In: 2010 17th IEEE International Conference on Image Processing (ICIP). IEEE (2010)
6. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7), 1409–1422 (2012)
7. Kalal, Z., Mikolajczyk, K., Matas, J.: Forward-backward error: Automatic detection of tracking failures. In: 2010 20th International Conference on Pattern Recognition (ICPR). IEEE (2010)
8. Kalal, Z., Matas, J., Mikolajczyk, K.: Online learning of robust object detectors during unstable tracking. In: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops). IEEE (2009)
9. Kalal, Z., Matas, J., Mikolajczyk, K.: P-N learning: Bootstrapping binary classifiers by structural constraints. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2010)
10. Shan, C., Tan, T., Wei, Y.: Real-time hand tracking using a mean shift embedded particle filter. Pattern Recognition 40(7), 1958–1970 (2007)
11. Kolsch, M., Turk, M.: Fast 2d hand tracking with flocks of features and multi-cue integration. In: Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2004). IEEE (2004)

A New Heuristic Algorithm for Improving Total Completion Time in Grid Computing

Soheil Anousha[1,*], Shoeib Anousha[1], and Mahmood Ahmadi[2]

[1] Department of Computer Engineering, Payame Noor University, Tehran, Iran
{Soheil.anousha,Shoeib.anousha}@gmail.com
[2] Department of Computer Engineering, University of Razi, Kermanshah, Iran
M.ahmadi@razi.ac.ir

Abstract. Grid is a large scale distributed system, concerned with coordinated resource sharing and problem solving. To make effective use of the huge capabilities of the grids, efficient task scheduling heuristic for improving time are required. Total completion time is an important factor for assessment of algorithm in Grid. Many algorithms for decreasing of total completion time have been implemented so far. In this paper, we propose new scheduling heuristic based on well known task scheduling algorithms, Min-Min. The proposed heuristic tries to use the advantages of this basic algorithm and avoids its fails. To achieve this, the proposed heuristic presented a new strategy for choosing appropriate resource in any round. The simulation results by Gridsim show that the new heuristic can improve performance for a variety of assumptions.

Keywords: Grid, resource, task scheduling, heuristic, Min-Min, completion time.

1 Introduction

Several definitions have been presented for grid [2, 3], but a standard definition for the grid with respect to all directions, is as follows. "Grid is a decentralized and parallel system that provides the resource sharing of costs, accessibility and quality of services." To make effective use of the huge capabilities of the computational grids, efficient task scheduling algorithms are required [9]. Many Grid task scheduling algorithms such as [15, 16] have some features in common, that are performed in multiple steps to improve quality of service. The well known example of algorithms is Min-Min [17]. This algorithm estimate completion times of each of the tasks on each of the grid resources. The Min-Min seems to do worse operation, whenever the number of small tasks is much more than the large ones. So, proposing a new algorithm to resolve the above mentioned problem is required. This paper offers a new task scheduling algorithm to resolve this problem with applying the Min-Min or Max-Min to schedule. To select the algorithm for first scheduling, we propose new Makespan. The most important of factor that can be improved by our algorithm is total completion

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_3, © Springer-Verlag Berlin Heidelberg 2014

time. The remainder of this paper is organized as follows. Related works are presented in section 2. In section 3, existing task scheduling algorithms is presented. In section 4, a new scheduling algorithm is proposed. In section 5, the experimental results are presented. Finally, section 6 concludes the paper and presents future works.

2 Related Works

We need to propose new scheduling algorithm for optimal use of available resources in the network and getting the less total completion time. These algorithms like [8, 9, 10 and 11] have some advantages and disadvantages. Our proposed algorithm tries to select best strategy and make a good balance. For achieving this, our algorithm proposes new strategy for selecting best task for scheduling in any round.

3 Existing Task Scheduling Algorithms

Generally, the scheduling algorithms are divided into two basic categories: **immediate mode** scheduling and **batch mode** scheduling [9]. In Immediate mode task is mapped onto a resource as soon as it arrives at the scheduler. For this mode we can mentioned MET and MCT. The ***MET*** (minimum execution time) heuristic assigns each task to the machine that performs that task's computation in the least amount of execution time [17]. The ***MCT*** (minimum completion time) heuristic assigns each task to the machine so that the task will have the earliest completion time [17]. ***MET*** and ***MCT*** deployed in SmartNet [6] and have ***O(R)*** time complexity when we have ***R*** resources. Another heuristic that do not use the execution or completion time of the tasks is OLB. OLB (opportunistic load balancing) assigns each task to the resource that becomes ready next, without considering the execution time of the task on that resource. When more than one resource becomes ready, one resource is arbitrarily chosen. In the implementation considered in [7], it takes ***O(R)*** time to find the assignment when we have ***R*** resources. One advantage of OLB is its simplicity, but because OLB does not consider expected task execution times, the mappings it finds can result in very poor Makespans [5]. In the batch mode, tasks are not mapped onto the resources as they arrive; instead they are collected into a set that is examined for mapping at prescheduled times called mapping events. The independent set of tasks which is considered for mapping at the mapping events is called a meta-task [14]. Min-Min, Max-Min and Sufferage are examples of this type.

3.1 Sufferage Algorithm

Suffrage algorithm is based on the idea that a task should be assigned to a certain resource and if it does not go to that resource, the most it will suffer [5]. Suffrage value for each task is defined through the following equation:

$$SV = secondMCT - MCT \quad (1)$$

where **SV** is the suffrage value and ***MCT*** denotes the minimum completion time and ***secondMCT*** denotes the second minimum completion time. Tasks with high suffrage values take precedence. Suffrage heuristic takes ***O(WRT)***, where $\boldsymbol{1 \leq W \leq T}$. In the worst case ***W*** is equal to ***T*** and in the best case ***W*** is equal to ***1*** [10].

3.2 Min-Min and Max-Min Algorithm

Min-Min algorithm starts with a set of all unmapped tasks. The machine that has the minimum completion time for all jobs is selected. Then the job with the overall minimum completion time is selected and mapped to that resource. The ready time of the resource is updated. This process is repeated until all the unmapped tasks are assigned. Compared to ***MCT*** this algorithm considers all jobs at a time. Time complexity of Min-Min algorithm when we have **R** resources and ***T*** tasks is $\boldsymbol{O(T^2R)}$. Max-Min is very similar to Min-Min algorithm. Like the Min-Min, the machine that has the minimum completion time for all jobs is selected. Then unlike the Min-Min, the job with the overall maximum completion time is selected and mapped to that resource. The ready time of the resource is updated. This process is repeated until all the unmapped tasks are assigned. The idea of this algorithm is to reduce the wait time of the large jobs. This algorithm takes $\boldsymbol{O(T^2R)}$ time, when we have ***R*** resources and ***T*** tasks. The Min-Min and Max-Min algorithm are given in Figure 1. For Max-Min algorithm, in line (7) of Figure 1, ***"minimum"*** would be changed to ***"maximum"***. As shown in Figure1, firstly it computes the amount of task completion time $\boldsymbol{CT_{ij}}$ for all tasks in **MT** on all resources from the following equation:

$$CT_{ij} = ET_{ij} + r_j \tag{2}$$

$\boldsymbol{CT_{ij}}$ is completion time and $\boldsymbol{ET_{ij}}$ is expected execution time of task ***i*th** on resource ***j*th**, and $\boldsymbol{r_j}$ is the ready time for resource ***j*th** ($\mathbf{r_j}$ is the ready time or availability time of resource **j** after completing previously assigned jobs). After that, the set of minimum expected completion time for each task in **MT** is found **(resource discovery)**, then the task with the overall minimum expected completion time from ***MT*** is selected and assigned to the corresponding resource **(resource selection)**.

(1) for all tasks t_i in MT
(2) for all machines mj
(3) $CT_{ij} = ET_{ij} + r_j$
(4) do until all tasks in MT are mapped
(5) for each task t_i in MT
(6) Find minimum CT_{ij} and resource that obtains it.
(7) Find the task t_k with the minimum CT_{ij}.
(8) Assign t_k to resource m_l that
(9) Delete t_k from MT
(10) Update r_l
(11) Update CT_{ij} for all i
(12) End do.

Fig. 1. The pseudo code of Min-Min (Max-Min) algorithm

4 The Proposed Heuristic

Our proposed scheduling heuristic is presented in Figure 2. Firstly this algorithm like the Min-Min algorithm, computes minimum completion time of all tasks on available resources. After that, the resource according to the appropriate Makespan should be chosen. The Makespan of the proposed algorithm computes as follows:

$$M(R_i)= [\text{Total value of all possibilities tasks except maximum value}]^{0.5} \quad (2)$$

(***R_i*** is the ***ith*** resource in our resources) It means that the Makespan could be achieved from root of summation times in all tasks except of maximum value in any resource. Also MVPT is an another parameter that described as follows:

$$MVPT= [\text{Maximum value of possibilities Tasks}] \quad (3)$$

After the computation of the Makespan and MVPT, we compare two parameters. If MVPT value is bigger than Makespan, it means that there exists a few long tasks along with too many short tasks, so it is better that we select Max-Min algorithm because Max-Min algorithm outperforms Min-Min, in otherwise if MVPT value is equal or less than Makespan, it shows that the number of long tasks is more than the number of short tasks, so the case where Min-Min outperforms Max-Min and we must select Min-Min algorithm. This procedure continued until all tasks on MT mapped. So the proposed algorithm chooses appropriate task in any round.

4.1 Time Complexity of Proposed Heuristic

The order of this algorithm is depending to two loop that mentioned in line (4) and (5) and also it's should be operable on all tasks (line (2)). So, this algorithm, likes the Min-Min takes ***$O(T^2R)$*** time, when we have ***R*** resources and ***T*** tasks.

4.2 Flowchart of Proposed Heuristic

Figure 3 shows the flowchart of our scheduling heuristic. This flowchart shows the procedure of our scheduling heuristic in selecting appropriate task for scheduling in any round.

4.3 An Illustrative Example

As a simple example, assume there is a grid environment with two resources. The completion time of the tasks are depicted in Table 1.

```
for all tasks ti in MT
    for all machines mj
        CTij = ETij + rj
do until all tasks in MT are mapped
    for each task ti in MT
  Find minimum CTij and resource that obtains it.
    If MVPT > Makespan(Ri)
        Find the task tk with the Maximum value of CT
        Assign tk to resource ml that
    Delete tk from MT
    Update rl
    Update CTij for all i
Else
        Find the task tk with the Minimum value of CT
        Assign tk to resource ml that
    Delete tk from MT
    Update rl
    Update CTij for all i
End do.
```

Fig. 2. Proposed heuristic

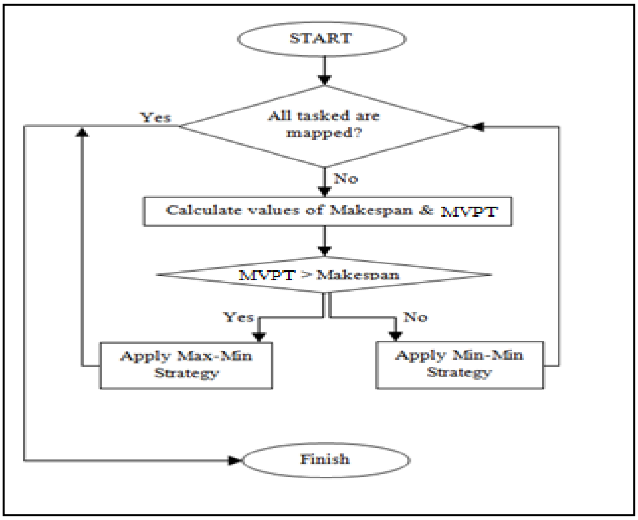


Fig. 3. Flowchart of proposed heuristic

Table 1. Completion time of the tasks on each of the resources

Tasks / Resource	R_1	R_2	R_3
T_1	3	5	19
T_2	4	6	20
T_3	6	13	14
T_4	9	14	29

Figure 4 includes two Gantt charts representing the results of applying Min-Min and Max-Min algorithms according to the values of completion time that described in Table 1. Also Figure 5 shows Gantt charts of our algorithm. Comparing the two figures shows that the proposed algorithm could obtains a better time unlike the Min-Min and Max-Min. Also the proposed algorithm uses Resource 3 and helps load balancing. Our algorithm computes the Makespan.

$$\begin{cases} \text{For } R_1: & 9 > (3+4+6)^{0.5} \\ \text{For } R_2: & 14 > (13+6+5)^{0.5} \\ \text{For } R_3: & 29 > (20+15+18)^{0.5} \end{cases}$$

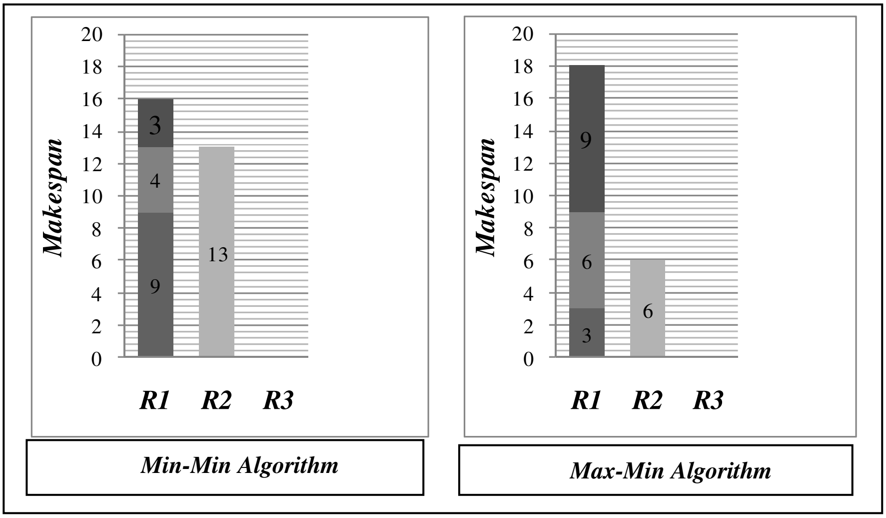


Fig. 4. Gantt chart of Min-Min and Max-Min algorithm

So, we use the Max-Min strategy for first scheduling. This shows that when we have a resource that have value greater than other resource, it's better that we use Max-Min strategy for first scheduling.

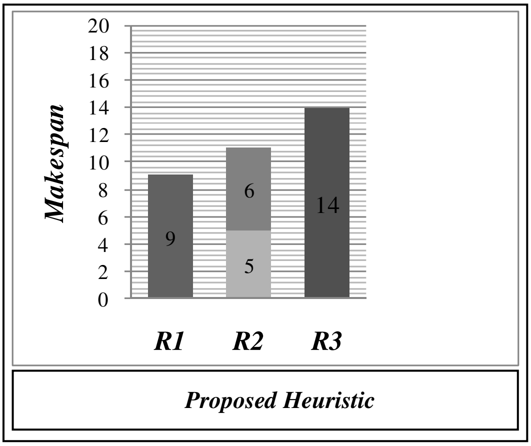


Fig. 5. Gantt chart of proposed heuristic

5 Simulation and Experimental Results

To compare and evaluate the proposed algorithm with other algorithms such as Max-min and Min-min a simulation environment known as GridSim toolkit [13] has been used. Our experimental testing performed in three assumptions:

1. **Assumption I:** A few short tasks along with many long tasks; i.e. the case where Min-Min outperforms Max-Min.

2. **Assumption II:** A few long tasks along with many short tasks; i.e. the case where Max-Min outperforms Min-Min.

3. **Assumption III:** With random tasks.

Number of resources is chosen to be 5. Two different numbers of tasks has been chosen: 200 for light and 1000 for medium load. Result of this simulation as follows: In Figure 6, which Min-Min algorithm outperforms Max-Min, the proposed algorithm acts like Min-Min.

In Figure 7, which Max-Min outperforms Min-Min, the proposed algorithm acts like Max-Min. Finally in Figure 8, with random tasks, our proposed algorithm outperforms Min-Min and Max-Min. It means that proposed algorithm in any assumption tries to improve total completion time.

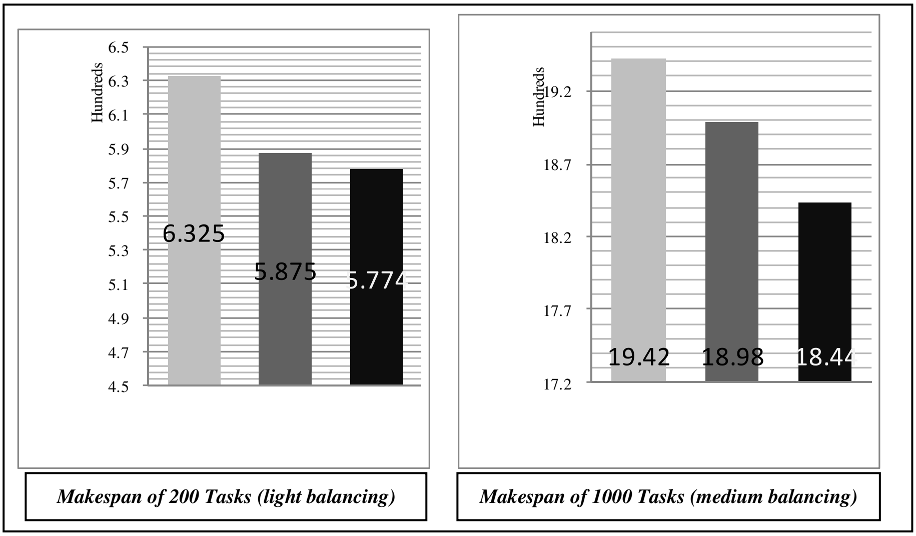


Fig. 6. Assumption I

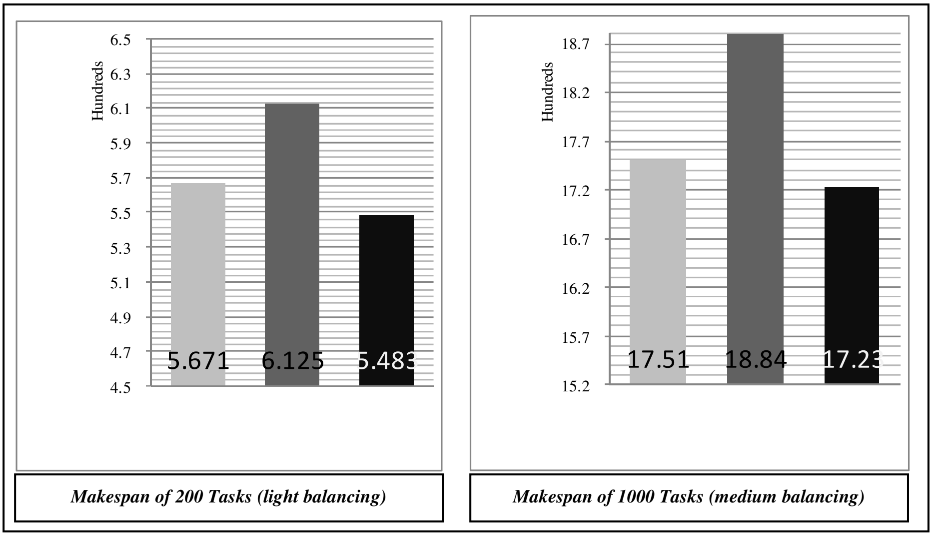


Fig. 7. Assumption II

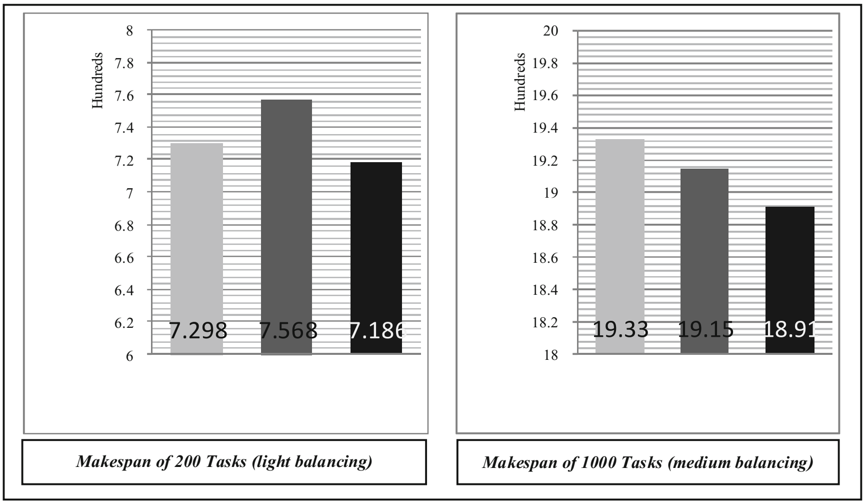


Fig. 8. Assumption III

6 Conclusion and Future Works

To overcome the limitations of the Min-Min algorithm, in this paper a new task scheduling heuristic based on Min-Min was presented. This heuristic algorithm proposed a new Makespan to select the resource for scheduling and uses the advantages of Min-Min algorithm and covers their disadvantages. The experimental results obtained by applying our algorithm within the GridSim simulator, shows that the proposed algorithm is outperforms better or minimal like Min-Min in low or medium load. This study concerned task execution time. For future works, we can apply other issues like load balancing.

References

[1] Foster, I.: Globus Toolkit Version 4: Software for serviceoriented systems. In: Jin, H., Reed, D., Jiang, W. (eds.) NPC 2005. LNCS, vol. 3779, pp. 2–13. Springer, Heidelberg (2005)

[2] Litzkow, M., Livny, M., Mutka, M.: Condor - A Hunter of Idle Workstations. In: Proceedings of the 8th International Conference of Distributed Computing Systems, pp. 104–111 (June 1988)

[3] Foster, I., Kesselman, C.: The Grid: Blueprint for a future computing Infrastructure. Morgan Kaufmann Publishers, USA (1999)

[4] Yagoubi, B., Slimani, Y.: Task Load Balancing Strategy for Grid Computing. Journal of Computer Science 3(3), 186–194 (2007)

[5] Maheswaran, M., Ali, S., Jay Siegel, H., Hensgen, D., Freund, R.F.: Dynamic Mapping of a Class of Independent Tasks onto Heterogeneous Computing Systems. Journal of Parallel and Distributed Computing 59, 107–131 (1999)

[6] Freund, R.F., Gherrity, M., Ambrosius, S., Campbell, M., Halderman, M., Hensgen, D., Keith, E., Kidd, T., Kussow, M., Lima, J.D., Mirabile, F., Moore, L., Rust, B., Siegel, H.J.: Scheduling Resource in Multi-User, Heterogeneous, Computing Environment with SmartNet. In: The Proceeding of the Seventh Heterogeneous Computing Workshop (1998)

[7] Braun, T.D., Jay Siegel, H., Beck, N., Boloni, L.L., Maheswaran, M., Reuther, A.I., Robertson, J.P., Theys, M.D., Yao, B.: A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems. Journal of Parallel and Distributed Computing 61, 810–837 (2001)

[8] Dong, F., Luo, J., Gao, L., Ge, L.: A Grid Task Scheduling Algorithm Based on QoS Priority Grouping. In: The Proceedings of the Fifth International Conference on Grid and Cooperative Computing (GCC 2006). IEEE (2006)

[9] Parsa, S., Entezari-Maleki, R.: RASA: A New Grid Task Scheduling Algorithm. International Journal of Digital Content Technology and its Applications 3(4) (December 2009)

[10] Etminani, K., Naghibzadeh, M.: A Min-min Max-min Selective Algorithm for Grid Task Scheduling. In: The Third IEEE/IFIP International Conference on Internet, Uzbekistan (2007)

[11] Afzal, A., McGough, A.S., Darlington, J.: Capacity planning and scheduling in Grid computing environment. Journal of Future Generation Computer Systems 24, 404–414 (2008)

[12] Brucker, P.: Scheduling Algorithms, 5th edn. Springer (2007)

[13] Buyya, R., Murshed, M.: GridSim: A toolkit for the odeling and simulation of distributed resource management and scheduling for grid computing. Journal of Concurrency and Computation Practice and Experience, 1175–1220 (2002)

[14] Benjamin Khoo, B.T., Veeravalli, B., Hung, T., Simon See, C.W.: A multi-dimensional scheduling scheme in a Grid computing environment. Journal of Parallel and Distributed Computing 67, 659–673 (2007)

[15] Czajkowski, K., Foster, I., Karonis, N., Kesselman, C., Martin, S., Smith, W., Tuecke, S.: A resource management architecture for metacomputing systems. In: Feitelson, D.G., Rudolph, L. (eds.) JSSPP 1998. LNCS, vol. 1459, pp. 62–82. Springer, Heidelberg (1998)

[16] Litzkow, M., Livny, M., Mutka, M.: Condor – A Hunter of Idle Workstations. In: Proc. the 8th International Conference of Distributed Computing Systems, San Jose, California, pp. 104–111 (June 1988)

[17] He, X., Sun, X.-H., Laszewski, G.V.: QoS Guided Min-min Heuristic for Grid Task Scheduling. Journal of Computer Science and Technology 18, 442–451 (2003)

[18] Izakian, H., Abraham, A., Snasel, V.: Comparison of heuristics for scheduling independent tasks on heterogeneous distributed environments. In: CSO 2009: Proceedings of the 2009 International Joint Conference on Computational Sciences and Optimization, pp. 8–12. IEEE Computer Society, Washington, DC (2009)

[19] Xhafa, F., Abraham, A.: Computational models and heuristics methods for grid scheduling problems. Future Generation Computer Systems 26, 608–621 (2010)

[20] Choon Lee, Y., Subrata, R., Zomaya, A.Y.: On the Performance of a Dual-Objective Optimization Model for Workflow Applications on Grid Platforms. IEEE Transactions on Parallel and Distributed Systems 20(9), 1273–1284 (2009)

[21] Priya, S.B., Subramaniam, C., Ravichandran, T.: On Demand Check Pointing for Grid Application Reliability using Communicating Process Model. IEEE (2011)

[22] Kadhim, S.J., Al-Aubidy, K.M.: Design and Evaluation of a Fuzzy-Based CPU Scheduling Algorithm. In: Das, V.V., et al. (eds.) BAIP 2010. CCIS, vol. 70, pp. 45–52. Springer, Heidelberg (2010)

Parallel Processing of Multimedia Data in a Heterogeneous Computing Environment

Heegon Kim, Sungju Lee, Yongwha Chung, Daihee Park, and Taewoong Jeon*

Dept. of Computer and Information Science,
Korea University, Sejong, Korea
{khg86,peacfeel,ychungy,dhpark,jeon}@korea.ac.kr

Abstract. Recently, many multimedia applications can be parallelized by using multicore platforms such as CPU and GPU. In this paper, we propose a parallel processing approach for a multimedia application by using both CPU and GPU. Instead of distributing the parallelizable workload to either CPU or GPU(*i.e.*, homogeneous computing), we distribute the workload simultaneously into both CPU and GPU(*i.e.*, heterogeneous computing) by using OpenCL. Based on the experimental results with a photomosaic application, we confirm that the proposed parallel processing approach can provide better performance than the typical parallel processing approach by utilizing the given resource maximally.

Keywords: CPU, GPU, Heterogeneous Computing, OpenCL.

1 Introduction

As multicore processors are used for handheld devices as well as PCs/servers, parallel processing approaches have been developed for many applications[1-2]. For example, many approaches have been reported to parallelize multimedia applications [3-4]. Furthermore, many users create their own content using these devices as handheld devices such as smartphones become powerful.

In this paper, we focus on parallelizing multimedia applications by using both CPU and GPU. In fact, these applications have sufficient parallelism, and many parallel processing results have been reported[5-7] by using general-purpose programming on GPU such as Nvidia's CUDA[8], in addition to Pthread[9] on CPU. Recently, OpenCL[10] has been defined as a standard for heterogeneous parallel computing. It provides a cross-platform framework for writing software able to run on different kinds of devices, from multicore CPUs to GPUs. That is, a parallel program written with OpenCL can be executed on either CPU or GPU[11]. Generally, it is true that GPU can provide better performance than CPU for multimedia applications. However, a current multicore CPU is also a powerful processor, and thus, when used together with GPU, can reduce the total execution time.

We propose a load balancing approach which can overcome the performance limit of either CPU-only or GPU-only execution. We first parallelize a given multimedia

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_4, © Springer-Verlag Berlin Heidelberg 2014

application with OpenCL, and measure its execution time on CPU and GPU, respectively. Then, we partition the parallelized workload into two parts, based on the relative performance of GPU over CPU. Finally, we assign the GPU-portion of workload to GPU by using a non-blocking command, and then assign the remaining parallel portion to CPU without waiting for a result from GPU. By reducing the idle time on either CPU or GPU, we overlap the GPU execution maximally with the CPU execution.

The rest of the paper is structured as follows. Section 2 explains OpenCL[10] and multimedia application *Photomosaic*[12]. Section 3 describes our proposed load balancing approach. The experimental results are given in Section 4, and conclusions are provided in Section 5.

2 Background

2.1 OpenCL

OpenCL[10] is an open standard aimed at providing a programming environment suitable to access heterogeneous architectures. In particular, OpenCL(shown in Fig. 1) allows to execute computational workloads on various multicore processors. Considering the increasing availability of such types of processors, OpenCL is playing a crucial role in enabling portable applications to access a wide range of computational resources. To achieve this aim, various levels of abstraction have been introduced in the OpenCL model.

- *Platform* performs an abstraction of the number and type of computing devices in a hardware platform. At this level are made available to developers the routines to query and to manage the computing devices, to create the contexts and work queues for submission of sets of instructions called **kernels**.
- *Execution* is based on the concept of kernel which is a collection of instructions executed on the computing device, multicore CPU or GPU, called OpenCL **device**. An OpenCL application can be divided in two programs: **host** and **kernel**. The host program is executed on CPU. It defines the context for the kernels and manages their execution. Especially, when a kernel is submitted for execution by the host, an **index space** is defined. An instance of the kernel executes for each point in this index space. This kernel instance is called a **work-item** and is identified by its point in the index space, which provides a global ID for the work-item. Each work-item executes the same code on distinguished data. That is, work-items are organized into work-groups providing a more coarse-grained decomposition of the index space.
- *Language* describes the syntax and programming interface for writing kernels(set of instructions that execute on computing device such as multicore CPUs or GPUs).

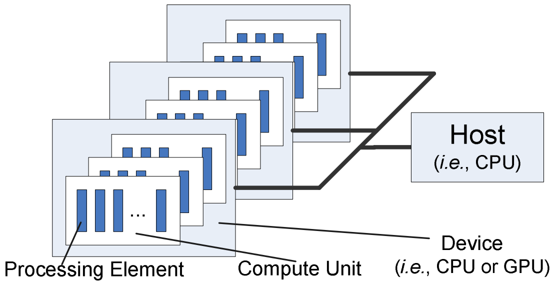


Fig. 1. Platform model of OpenCL

2.2 Photomosaic

A photomosaic[12] is a compound word of "Photograph" and "Mosaic". The photomosaic divides a large image into several small parts, which are converted into small tile images of similar colors.

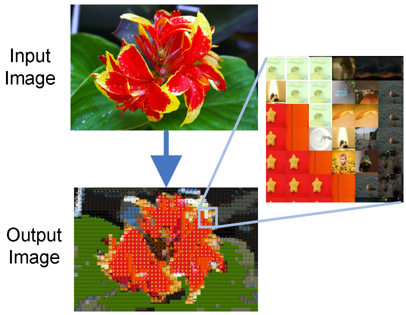


Fig. 2. Result of the photomosaic

Fig. 2 shows the similarity between the original image and the result image of the photomosaic. The result image is composed of many smaller tile images. In this paper, the photomosaic iterates the loop of image conversion 5 times per pixel.

3 Parallel Photomosaic

The performance of each core of CPU is better than GPU's, whereas the number of CPU cores is less than the number of GPU cores. The GPU which has hundreds of cores is more advantageous, if calculation is made of a lot of iterations of the same operation. A large number of studies of GPU-equipped environments using only the GPU parallel processing have been published[13-14].

The photomosaic does not have data dependency among the tile images. Therefore, a parallel photomosaic by OpenCL is processed using compute units for each tile images. The host program is waiting during the execution of the kernel function, because typical OpenCL programs are performed by synchronization using *blocking mode*(see Fig. 3).

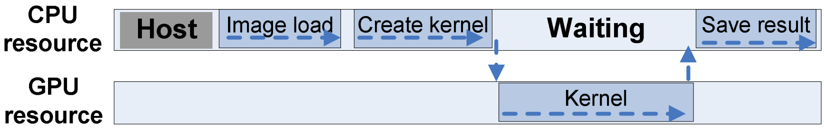


Fig. 3. Typical parallel processing of photomosaic using GPU

In heterogeneous computing environments, we propose an approach which improves performance using not only GPU but also CPU to reduce the CPU idle time(*i.e.*, waiting time). OpenCL allows asynchronous processing using *non-blocking mode*. In this paper, non-blocking mode is used in order to reduce the CPU idle time. In non-blocking mode, both CPU and GPU resources can be used simultaneously as shown in Fig. 4. Since the idle time is reduced, the proposed approach can effect a speedup higher than can be achieved by typical parallel processing.

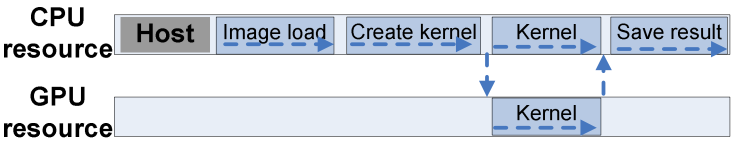


Fig. 4. Proposed parallel processing of photomosaic using both GPU and CPU

4 Experimental Results

For evaluating the proposed approach, we used AMD Phenom II X4 955 Processor, GeForce GTX 285, and the target image with 3072×2048 resolution. The number of tile images is 1000. AMD Phenom II X4 955 Processor has four cores, and GeForce GTX 285 has 240 cores. However, the GPU core provides lower performance than the CPU core. Also, many typical parallel processing studies with GPU have focused on GPU only.

First, the execution time of the photomosaic was measured for evaluating parallel OpenCL speedup. The photomosaic was measured in three ways: sequential, parallel using GPU-only by OpenCL, and parallel using multicore CPU-only by OpenCL. Table 1 shows the sequential and parallel execution times of the photomosaic application. "Multicore CPU-only" was measured using multicore CPU, and "GPU-only" was measured using GPU. "Multicore CPU(x%)+GPU(y%)" was measured using both multicore CPU and GPU, and multicore CPU had x% portion while GPU had y% portion. The result shows that the performance of using CPU-only by OpenCL provides super speedup(*i.e.*, a 4-core CPU has a speedup of 17). The reason is that the cache-hit ratio was highly improved with the increased number of cores.

Table 1. Sequential and parallel execution times of the photomosaic

		Execution time(sec)
Sequential processing		340.62
Parallel processing	Multicore CPU-only	20.29
	GPU-only	10.53
	Multicore CPU(50%) + GPU(50%)	12.07
	Multicore CPU(25%) + GPU(75%)	8.40

Next, the execution time of the photomosaic with the workload divided into two parts was measured, in which one part was performed by CPU and the other part was performed by GPU. As Table 1 shows, the photomosaic that was divided into 25% CPU portion and 75% GPU portion can provide better performance than the one using multicore CPU-only or GPU-only. These portions were constrained by index space, therefore the division into two parts is not possible in certain proportions depending on the GPU performance and CPU performance(*i.e.*, CPU(33%) + GPU(66%)). The proposed approach can have a speedup of 40, and can yield 25% better performance than the one using GPU-only by OpenCL. However, if the 2-part division is made inappropriately, the proposed approach provides lower performance than GPU-only. Fig. 5 shows the speedups with OpenCL achieved by four different ways of parallel processing.

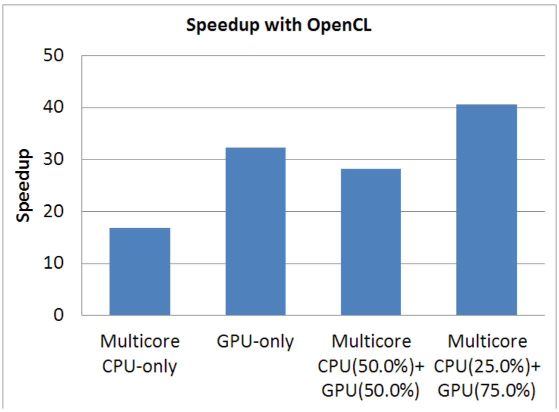


Fig. 5. Speedup with OpenCL

5 Conclusions

We have proposed an efficient heterogeneous parallel processing approach to reduce CPU idle time. The approach, which uses both CPU and GPU by OpenCL, decreases total execution time for better performance.

Experiments with the use of both CPU and GPU for parallel processing have demonstrated that our parallel processing approach can provide a speedup of 40 and (if properly load-balanced between CPU and GPU) 25% better performance than the generally used parallel approach using GPU only.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(funded by the Ministry of Education, Science and Technology, 2012R1A1A2043679) and BK21 Plus Program.

References

1. Held, J., Bautista, J., Koehl, S.: From a Few Cores to Many: A Tera-Scale Computing Research Overview. Intel White Paper (2006)
2. Levy, M., Conte, T.: Embedded Multicore Processors and Systems. IEEE Micro 29, 7–9 (2009)
3. Sihn, K., Baik, H., Kim, J., Bae, S., Song, J.: Novel Approaches to Parallel H.264 Decoder on Symmetric Multicore Systems. In: Proc. of International Conference on Acoustics, Speech, and Signal Processing, pp. 2017–2020 (2009)
4. Chen, W., Hang, H.: H.264/AVC Motion Estimation Implementation on CUDA. In: Proc. of International Multimedia and Expo Conf., pp. 697–700 (2008)
5. Shams, R., Sadeghi, P., Kennedy, R., Hartley, R.: A Survey of Medical Image Registration on Multicore and the GPU. IEEE Signal Processing Magazine 27(2), 50–60 (2010)
6. Bienia, C., Kumar, S., Singh, J., Li, K.: The PARSEC Benchmark Suite: Characterization and Architectural Implications. In: Proc. of International Conference on Parallel Architectures and Compilation Techniques, pp. 72–81 (2008)
7. Kim, H., Lee, S., Chung, Y., Pan, S.: Parallelizing H.264 and AES Collectively. KSII Tr. Internet & Info. Systems 7(9), 2326–2337 (2013)
8. NVidia, NVidia CUDA Compute Unified Device Architecture Programming Guide, NVidia (2008)
9. Akhter, S., Roberts, J.: Multi-Core Programming - Increasing Performance through Software Multi-Threading. Intel Press, Hillsboro (2006)
10. Stone, J., Gohara, D., Shi, G.: OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems. Computing in Science and Engineering 12(3), 66–73 (2010)
11. Gaetano, R., Pesquet-Popescu, B.: OpenCL Implementation of Motion Estimation for Cloud Video Processing. In: Proc. of International Symposium on Multimedia Signal Processing, pp. 1–6 (2011)
12. Silvers, R., Hawley, M.: Photomosaics. Henry Holt, New York (1997)
13. Cao, J., Xie, X.-f., Liang, J., Li, D.-d.: GPU Accelerated Target Tracking Method. In: Jin, D., Lin, S. (eds.) Advances in MSEC Vol. 1. AISC, vol. 128, pp. 251–257. Springer, Heidelberg (2011)
14. Davendra, D., Zelinka, I.: GPU Based Enhanced Differential Evolution Algorithm: A Comparison between CUDA and OpenCL. Intelligent Systems Reference Library, vol. 38, pp. 845–867 (2013)

Optimization Problems Related to Hamiltonian Paths

Sylvia Encheva

Stord/Haugesund University College, Bjørnsong 45, 5528 Haugesund, Norway
sbe@hsh.no

Abstract. Rapid development of new technologies minimizes the significance of distances but increases complexity levels of previously known Hamiltonian paths problems by introducing additional variables. We are employing application of lattices and weight functions for decreasing the amount of choices to be evaluated in the process of looking for an optimal solution.

Keywords: New technologies, Weight functions, Optimization.

1 Introduction

Hamiltonian paths are well researched area and have been of interest to scientists from a number of different fields, [7]. Hamiltonian paths were originally introduced in order to optimize postmen's' routes, [11]. One of the implicit assumptions following such optimization problems is that all vertices of a graph under investigation are of equal importance. The travelling salesmen problem is another term regarding similar problems. However, the latter formulations open for inclusion of other parameters than distance, [11]. Weighted graphs with edges of not necessarily equal importance are a step forward in addressing modern society issues, [6].

New technological developments minimize the significance of distances between vertices in a graph but increase complexity levels of previously known problems by introducing additional variables. We suggest focusing on vertices instead of edges. Application of lattices and weight functions is also discussed.

2 Weights

A Hamiltonian path or traceable path is a path that visits each vertex exactly once, [6] and [11]. Sufficient conditions for a graph to possess Hamiltonian cycles and Hamiltonian paths are presented in [8]. Algorithms for Hamiltonian cycles are presented in [2] and [4]. Hamiltonian paths have been applied in the process of designing of web sites and distributed database systems [9].

The problem of aggregating a set of numerical readings in order to obtain a mean value is addressed in [10]. If $x_1, x_2, ..., x_n$ is a set of readings then the aggregating process is denoted as $Agg(x_1, x_2, ..., x_n)$. Addition of new elements with values

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_5, © Springer-Verlag Berlin Heidelberg 2014

greater than the current mean value results in an increased mean value and addition of new elements with values smaller than the current mean value results in an decreased mean value, [10]. Weights assigned to readings $x_1, x_2, ..., x_n$ must satisfy the following conditions

$$\sum_{j=1}^{n} w_{nj} = 1 \;\forall\; n \; and \; w_{nj} \geq 0 \;\forall\; n, j \cdot$$

For the case with n arguments $Agg(x_1,...,x_n) = \sum_{j=1}^{n} w_{nj} x_n$. The weights w_{nj} are uniquely determined by the formula

$$w_{ni} = w_{n1} \left(\frac{1}{w_{i,1}} - \frac{1}{w_{i-1,1}} \right), \; 2 \leq i \leq n$$

and the ratio $\frac{w_{n1}}{w_{n-1,1}}$ between the first element in the current iteration and the first element in the previous iteration, [10].

Calculation of weights w_{nj} for new elements can also be done by applying the Lukasiewicsz t-conorm $S(x_1,...,x_n)$, [5]. If factors like importances and frequencies are to be involved in the decision proses one may use the method of incorporating quantitative weights into aggregation, [1]. In statistics the weighted mean is presented as

$$W(x_1,...,x_n) = \frac{\sum_i n_i x_i}{\sum_i n_i}$$

where n_i are integer weights and x_i are frequencies of observations. W is further generalized to $W(x_1,...,x_n) = \sum_i w_i x_i, \; w_i \geq 0$ where the weight w_i corresponds to the i th input x_i.

3 Hamiltonian Paths

In this model vertices and edges are arranged in a semi distributive lattice. An optimal path is calculated by a new function.

Let this approach be illustrated with a comprehensive example. Consider a path with four vertices a, b, c, d. A semi distributive lattice is shown in Fig. 1. By ab we denote a path starting at a and ending at b and on each level of the lattice we introduce one additional element in each vertex. Without lost of generality we present paths starting at a only, since in practice an optimization problem of that

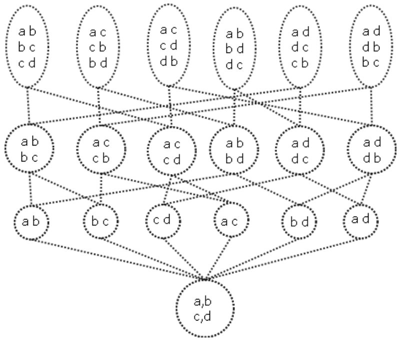


Fig. 1. A lattice related to a path with four vertices

kind starts at one particular vertex and we might as well call it a. A closer look at the second level of nodes in Fig. 1 indicates that they can be placed in sets with two nodes in each set where nodes in a set differ in one element only (keep in mind that we are working with undirected graphs).

Thus at the initial stage we have to choose between three vertices, i.e. b, c, d, (keep in mind that we have started with a). Once the second vertex is chosen, there are only two possible ways to continue, i.e.

if the second vertex is b, then ab, bc, cd or ab, bd, dc,
if the second vertex is c, then ac, cd, db or ac, cb, bd,
and
if the second vertex is d, then ad, dc, cb or ad, db, bc.

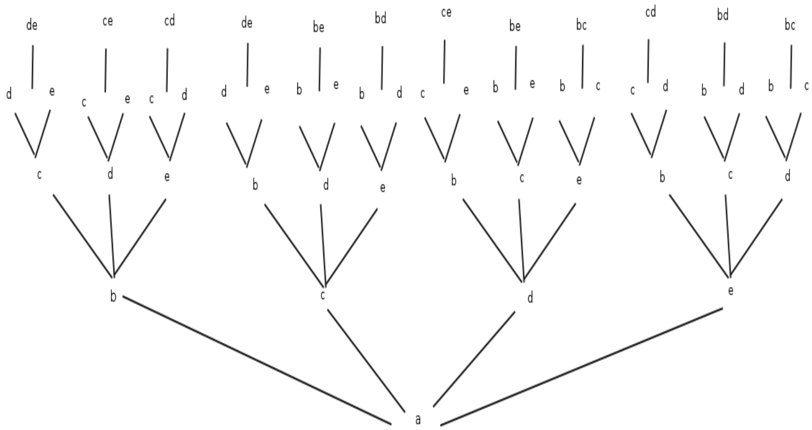


Fig. 2. A lattice related to a path with five vertices

The case with five vertices is presented in Fig. 2. It confirms the idea of reducing the number of options by grouping them into levels. An edge ending up with one letter is used when there are at least two different options, an edge ending up with two letters indicates that there are two options but they are identical with respect to distances.

Vertices represent geographical places. Their importance in terms of profit is changing. In order to address this we suggest inclusion of the following function $f(w_j) = \sum_{i,j=1}^{n} D_{ij} w_j$, where D_{ij} is the distance between vertices i and j and w_j is the importance of vertex j. The first leg of a journey requires $n-1$ calculations of the importance of each of the $n-1$ geographical places to chose from. On the next step there are $n-2$ choices but the calculations are ready due to the work done for the first choice. Continuing in the same fashion we complete the journey. Since our attempt is to model a world that is constantly changing we suggest all w_j to be considered as Yagers' weights.

4 Conclusion

This work presents an idea for reducing computational complexities related to small sizes Hamiltonian paths involved in practical applications. Instead of considering the original graph theoretical problem in a holistic way we brake it into smaller chunks and let other variables influence our choice. We believe that similar approaches can be used in decision support systems.

References

1. Calvo, T., Mesiar, R., Yager, R.: Quantitative weights and aggregation. IEEE T. Fuzzy Systems 12(1), 62–69 (2004)
2. Chang, M.-S., Peng, S.-L., Liaw, J.-L.: Deferred query|an efficient approach for problems on interval and circular-arc graphs. In: Dehne, F., Sack, J.-R., Santoro, N., et al. (eds.) WADS 1993. LNCS, vol. 709, pp. 222–233. Springer, Heidelberg (1993)
3. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)
4. Deogun, J.S., Steiner, G.: Polynomial algorithms for Hamiltonian cycle in cocomparability graphs. SIAM J. Comput. 23, 520–552 (1994)
5. Dubois, D., Prade, H.: A review of fuzzy sets aggregation connectives. Information Science 36, 85–121 (1985)
6. Graham, R.L.: Handbook of Combinatorics. MIT Press (1995)
7. Gross, J.L., Yellen, J.: Handbook of Graph Theory. CRC Press Inc. (2004)
8. Rahman, M.S., Kaykobad, M.: On Hamiltonian cycles and Hamiltonian paths. Information Processing Letters 94, 37–41 (2005)
9. Torres-Velazquez, R., Estivill-CastroSource, V.: Local Search for Hamiltonian Path with Applications to Clustering Visitation Paths. The Journal of the Operational Research Society 55(7), 737–748 (2004)

10. Yager, R., Rybalov, A.: Noncommutative self-identity aggregation. Fuzzy Sets and Systems 85, 73–82 (1997)
11. West, D.B.: Introduction to Graph Theory. Prentice-Hall, Englewood Cliffs (2001)
12. Wille, R.: Concept lattices and conceptual knowledge systems. Computers Math. Applications 23(6-9), 493–515 (1992)

Selecting Processes Supported by Fuzzy Calculations

Sylvia Encheva

Stord/Haugesund University College,
Bjørnsong 45, 5528 Haugesund, Norway
sbe@hsh.no

Abstract. Cooperative decision making is a well discussed research topic and has been often practised in a form of majority voting. Nowadays complicated life situations which involve a large number of conditional attributes and several alternative solutions require new, efficient and non the less systematic ways of handling cooperative decision making. In this paper we propose an approach for automated selection of a set of options being similar at least to a beforehand chosen degree. Fuzzy calculations are applied for accommodating non binary evaluations.

Keywords: Fuzzy calculations, Cooperative decision making, Similarities.

1 Introduction

Cooperative decision making has been studied from many different prospectives. In [10] the conditions under which the process for arriving at a cooperative decision are first specified and then possible strategies chosen by corresponding parties are formulated. Democratic control on cooperative decision making has been discussed in [1]. The authors argue that 'democratic control does not guarantee optimal decisions for the cooperative' while 'suboptimal decisions result from simple majority rule'. More complicated cases involving several conditional attributes and a large number of alternatives require new approaches. Collaborative filtering describes techniques that use the known preferences of a group of users to predict the unknown preferences of a new user; recommendations for the new user are based on these predictions, [9]. Information obtained via fuzzy equivalence relations can be used to determine a limit for the degree of precision in which inputs should be measured, since a higher accuracy than the indistinguishability inherent to the fuzzy sets or fuzzy equivalence relations does not influence the resulting output of a fuzzy system, [7]. Similarity relations in fuzzy sets have been introduced in [12].

In this paper we propose an approach for automated selection of a set of options being similar at least to a beforehand chosen degree. Fuzzy calculations are applied for facilitating non binary evaluations. In addition we provide a solution for cases where different decision makers have different degrees of influence on the decision making.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_6, © Springer-Verlag Berlin Heidelberg 2014

2 Preliminaries

Let P be a non-empty ordered set. If $sup\{x,y\}$ and $inf\{x,y\}$ exist for all $x,y \in P$, then P is called a *lattice*, [2]. A *context* is a triple (G,M,I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively [2], [4], and [12]. Factorization of lattices by tolerance relations is discussed in [4].

For two fuzzy sets $\tilde{X}_1, \tilde{X}_2 \in L^U$, where L^U is the set of all L-fuzzy sets in U and $S(\tilde{X}_1,\tilde{X}_2) = \bigwedge_{x\in U}(\tilde{X}_1(x) \to \tilde{X}_2(x))$. $S(\tilde{X}_1,\tilde{X}_2)$ is the subsethood degree and expresses the truth value of "each element of $\tilde{X}_1$ is an element of $\tilde{X}_2$", [5]. A fuzzy concept lattice can be factorized by appropriate a-cut ${}^aE$ of the similarity E (note that ${}^aE = \{\langle c1, c2\rangle \mid a \le E(c_1,c_2)\}$ [8]), controlling thus the complexity by $a \in L$.

A partition $P = \{C_1, C_2, ..., C_n\}$ of a set E is a set of non-empty subsets of E, called its classes, pairwise disjoint and the union of which is equal to E. The relation defined on the set P_E of partitions of E by $P \le P'$ if any class of P is included in a class of P' is an order called refinement order. We say 'P is finer than P'', [3].

Cluster analysis partition data into sets (clusters) sharing common properties. A frequently used tool in cluster analysis is a dissimilarity function d on a set of objects E, measuring the degree of dissemblance between the elements in E. A dissimilarity on a set E is a function d from E^2 to the set R^+ or non-negative real numbers satisfying the following properties for all $e, e' \in E$: $d(e,e') = 0$ if and only if $e = e'$, $d(e,e') = d(e',e)$.

3 Decision Making

Selecting the most appropriate employ for a job requiring several skills is not always a straightforward procedure. Suppose several departments in one organization are to receive services from one person. A number of applicants for the vacant position are considered by the corresponding departments heads. Since humans usually prefer to use linguistic values the departments' heads agree in advance on which linguistic terms are to be used for assessing applicants' qualification levels.

In order to facilitate such cooperative decision making we propose application of fuzzy concepts. Each applicant's qualifications are evaluated by all departments heads. The outcome is then inserted in an information table. If the linguistic values are binary (f. ex. 'qualified' or 'not qualified'), one can build a corresponding concept lattice and then take the option placed in the top node or call for an interview all applicants placed in the top node. In case the position is not taken by any of the applicants placed in the top node of the concept lattice, one can consider applicants placed in a node just below the top node.

In a non binary case a fuzzy concept lattice has to be build instead. The problem with such lattices is that they very soon become too big, as a result of increased number of objects and attributes. In addition they also become quite difficult to read. Such obstacles can be avoided by applying factorization, [8].

Suppose six departments heads are evaluating twenty applicants applying five linguistic values. The outcome is inserted in an information table with twenty-one rows (one information row and twenty rows related to the twenty applicants, $A1, A2, ..., A20$) and seven columns (one information column and six columns related to each department, $D1, D2, ..., D6$). Subsequently, this information table is rewritten, where the linguistic values are replaced with numerical values, f. ex. very well qualified - (1), well qualified - (0.75), qualified - (0.5), somewhat qualified - (0.25), and not qualified - (0).

A fuzzy concept lattice is then build. In order to select the most suitable applicants, who will be invited for an interview, one applies factorization by f. ex. 0.75-cut. The block of tolerance containing the case $(1,1,1,1,1,1)$ is first to be considered. All applicants in this block may be called for an interview since their qualifications are similar at least in the degree 0.75 to the best possible one. If a 0.75-cut does not apply, i.e. there are no such applicants with such evaluations, one can consider a 0.5 -cut, f.ex. Naturally, the number of applicants is increasing when a lower degree of similarity is applied, [13].

The work so far assumes that all departments have equal amount of influence on the decision. In many cases it might be necessary to distinguish the degree of importance of such an employ for each department. This can be achieved by asking each department head to supply a coefficient of importance in a form of a number w_j such that $\sum_{j=1}^{6} w_j = 1, 0 \leq w_j \leq 1$. Then every row of the information table with numerical values $(c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}, c_{i6})$, will be multiplied by a vector $(w_1, w_2, w_3, w_4, w_5, w_6), 1 \leq i \leq 20$ using component wise multiplication, i.e. $(c_{i1} \cdot w_1, c_{i2} \cdot w_2, c_{i3} \cdot w_3, c_{i4} \cdot w_4, c_{i5} \cdot w_5, c_{i6} \cdot w_6)$, where c_{ij} is the qualification level of applicant $c_i, 1 \leq j \leq 20$ assigned by the head of the j-th department, $1 \leq j \leq 6$. The resulting information table is then used for building a fuzzy concept lattice and an appropriate factorization is performed afterwords.

Suppose four applicants, denoted by a, b, c, d in Fig. 1, are selected for final considerations due to the above described work. Applying refinement order on the set of four applicants we obtain partitions of the set placed in a lattice in Fig. 1. All applicants are placed in the top node of the lattice as a start. Nodes placed in the second row of the lattice illustrate how the four applicants can be placed in two sets according to a single attribute, it can be f. ex. age, sex, etc. Thus obtained subsets of applicants are further dived in smaller subsets according to another attribute, f. ex. experience, education, etc. This is particularly useful for further selection of an applicant or a subset of applicants possessing f. ex. a special skill or a combination of skills.

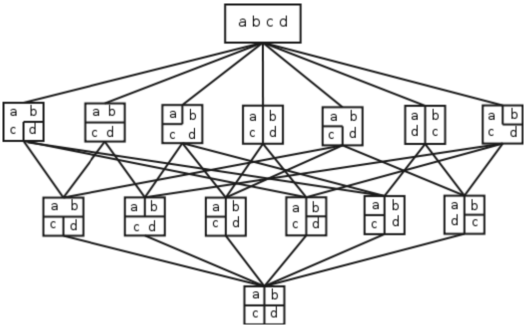


Fig. 1. Lattice of partitions

4 Conclusion

In this paper we address the problem of automated selection of optimal alternatives being similar at least with a given degree. The approach is based on the theory of fuzzy concept lattices and can accommodate attributes with different level of importance. The outcome does not just extract the best solution but presents also all other solutions that do not differ considerably from the desired one.

References

1. Anderson, B.L.: Democratic Control and Cooperative Decision Making: A Conceptual Framework. Journal of Agricultural Cooperation (1987)
2. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)
3. Caspard, N., Leclerc, B., Monjardet, B.: Finite Ordered Sets Concepts, Results and Uses. Cambridge University Press (2012)
4. Ganter, B., Wille, R.: Formal Concept Analysis. Springer (1999)
5. Goguen, J.: L-fuzzy sets. J. Math. Anal. Appl. 18, 145–174 (1967)
6. Klahr, D.A.: Computer Simulation of the Paradox of Voting. The American Political Science Review LX(2), 384–390 (1966)
7. Klawonna, F., Castrob, J.L.: Similarity in Fuzzy Reasoning. Mathware & Soft Computing 2, 197–228 (1995)
8. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic. Theory and Applications. Prentice Hall (1995)

9. Pennock, D.M., Horvitz, E.: Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach. In: IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, International Joint Conference on Artificial Intelligence (1999)
10. Rosing, J.: The Formation of Groups for Cooperative Decision Making Under Uncertainty. Econometrica 38(3), 430–448 (1970)
11. Wille, R.: Concept lattices and conceptual knowledge systems. Computers Math. Applications 23(6-9), 493–515 (1992)
12. Zadeh, L.A.: Similarity relations and fuzzy orderings. Information Science 3, 177–200 (1971)
13. Zadeh, L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90(2) (1997)

Formal Specification of Cyber Physical Systems: Case Studies Based on Hybrid Relation Calculus

Wenli Zhang, Lichen Zhang*, and Haibin Cai

Shanghai Key Laboratory of Trustworthy Computing
East China Normal University
Shanghai, China
zhanglichen1962@163.com

Abstract. Hybrid system is a dynamic mixture of continuous physical world and discrete control part. In hybrid systems, continuous and discrete dynamic behaviors are displayed by the continuous and discrete components of the system respectively. Due to the complicated mixture mechanism, it becomes very difficult for us to model such a system accurately and explicitly. To simplify the modeling process, Jifeng He proposes a hybrid parallel programming language. It is in the light of his intellectual work that we can model physical world as well as its interaction with the control program. This paper aims to apply the programming language proposed by Jifeng He to the specification of hybrid systems. To describe this language specifically, we used a simple system which consists of a motor boat on a lake modeled by the space R^2 .This example illustrates that using hybrid relation calculus to specify hybrid systems can make every subsystem less complicated and more detailed, so that we can get easy access to observe the physical mechanism of the system.

Keywords: hybrid systems, continuous and discrete, hybrid relation calculus, model.

1 Introduction

A Cyber physical system(CPS)[1] combines computation and physical process together, usually, also coupled with complicated clock interval[2] and perplexed communication mechanism[3]. In the terms of the behavior of the system event, CPS can be presented and described either in continuous model or in discrete program, and if you join the two components together, a hybrid system will appear with a dynamic mixture of continuous physical world and discrete control part. In hybrid systems, continuous and discrete dynamic behaviors are displayed by the continuous and discrete components of the system respectively. [4]During discrete dynamic transitions, system state varies in real-time and in discontinuous sequence while in continuous transitions, the system states make up a continuous function, and as the elapse of continuous time, they change in accordance with a differential equation. In fact, applications of hybrid systems[5]are everywhere to be seen, it plays an indispensable role in health care

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_7, © Springer-Verlag Berlin Heidelberg 2014

(the heart operation, the B-machine, proteins detecting) , aviation and aerospace, automotive manufacturing(the smart building, smart devices), digital chip designs and so on. Since hybrid system is critical to our daily life and industrial production, nothing will affect us as much as a detailed and specific description of such a system. However, due to the complicated mixture mechanism, it becomes very difficult for us to model hybrid systems[6]accurately and explicitly. To simplify the modeling process, Jifeng He proposes a hybrid parallel programming language[7] whose semantics are in line with the reality very well. It is in the light of his intellectual work that we can model physical world as well as its interaction with the control program.

Hybrid language[8] is adapted to large-scale logical systems that sit between modal and classic logic. There have been a lot of hybrid parallel programming languages put forward to model the complex systems so far. According to my survey, Kerstin Bauer once put forward a new modeling language called Hybrid Quartz in his PhD thesis[9].This new language is a bit similar with the language proposed by Jifeng He in that both of them have a precise formal semantics and make up for the gap between hybrid systems and their symbolic representation. Longfei Zhu[10] also has done some work related to it. Critical in his work are a series of algebraic laws[11] which can transform the physical process into a mathematical formula. No doubt due to his great work, program texts can be simplified and illustrated accurately and explicitly. And when it comes to the specification [12][13]of the hybrid systems, Rajeev Alur and Radu Grosu also have done much constructive work that the thesis Modular Specification of Hybrid Systems in Charon[14]by them introduces a new language named Charon for modular specification of interacting hybrid systems[15].

This paper aims to apply the programming language proposed by Jifeng He to the specification of hybrid systems. To describe this language specifically, we used a simple system which consists of a motor boat on a lake modeled by the space R^2. This example illustrates that using hybrid relation calculus to specify cyber physical systems can make every subsystem less complicated and more detailed, so that we can get easy access to observe the physical mechanism of the system.

2 A Hybrid Programming Language

The hybrid Programming Language proposed by Jifeng He in his paper "Hybrid Relation Calculus "extends the guarded common language, in which both the parallel composition and the sequential composition are included. **until** and **when** statements are also used to describe a completed physical event. Syntax of the language is listed as follows:

$$A::=\mathbf{skip}|\mathbf{chaos}|x:=e|\mathbf{delay}|!s$$

$$EQ::=F(v,v')=0|\mathbf{idle}|EQ||EQ$$

$$P::=A|P\sqcap P|P;P|P||P|P\lhd b\rhd P$$

$$EQ\ \mathbf{until}\ g|\mathbf{when}(G)|\mu X\bullet P(X)$$

$$G::=g\&P|G\square G$$

$$g::=s|g\cdot g|g+g$$

A stands for the alphabet of the program. **skip** does nothing. **chaos** indicates that the behaviors of the program is hardly predictable. Assignment statement x:=e is used to assign the value of expression e to discrete variable x instantaneously. It exhibits the change of discrete program. **delay** is similar to **skip** except that its stop time is unknown previously. !s is an output command that can emits the signal s.

F(v, v')=0 is an differential equation that is used to describe the continuous dynamics of variable v.

idle represents silent physical component, and only exhibits the passage of time. EQ1||EQ2 stands for their composite equation.

The notation PΠQ is defined as follows: PΠQ= $_{df}$ $P \vee Q$. P;P represents sequential composition and P||P defines parallel composition. Assume b is a Boolean expression, then

$P \lhd b \rhd Q =_{df} (b \wedge P) \vee (\neg b \wedge Q)$, *EQ* ***until*** *g* is a continuous statement that defines EQ when g is not triggered. ***when****(G)* is a guard statement concerning the input event guard

G is a set of guard. When g is triggered, P will execute immediately. If more than one event guards meet the trigger, the branch will be chosen at random.

For a signal s, its has two status: **absent** or **present**.

3 Case Study

To verify the hybrid programming language, a typical model concerning environment-controller interaction is built in which a motor boat and its sailing environment are included. The model is built up by three different components: the boat itself ,the water and the motor of the boat(the controller) . This is a cyber physical system where the motor controller is the computation component and the interaction of the boat and the water is the physical process. The function of the motor controller is guiding the boat to a specific point. For simplicity, we consider that the water have no boarders and the force of the wind. We define that if the distance between the destination and the position is less than infinitesimal number ε, the boat has arrived at the goal area. The water also has velocity, so the movement of the boat is affected not only by itself , but also by the flowing water. In this system, the motor controller is able to perceive the velocity of the motor but the situation of the outside environment(the velocity of the water) is completely unknown to the controller. This example is in a very typical scenario of hybrid system that every sub-module runs in a parallel way and they communicate with each other through shared variables and some complicated communication mechanism. We will analysis three parts of the system respectively as follows:

The Boat

The primary property of the boat is its velocity which is induced by two main factors: the speed generated by the operation of the engine and the water drift. The input variables include the velocity of the motor and the velocity of the water, and the output variable is the coordinates of the boat.

Table 1. Variables in Steam Boat-Controller System

Parameter	Value
motorX	the X direction of the motor's speed
motorY	the Y direction of the motor's speed
waterX	the X direction of the water's speed
waterY	the Y direction of the water's speed
boatX	the X direction of the boat's position
boatY	the Y direction of the boat's position
landingX	the X coordinate of the landing place
landingY	the Y coordinate of the landing place

The variables above are alphabets of the system. differential equation F(boatX, boatX')=0 specifies the continuous dynamics of variable boatX, where boatX' stands for the derivative of boatX, that is the velocity of the boat, so does F(boatY, boatY')=0. Let A be the input alphabet of the equation. F(boatX, boatX')=0 is used to present the position of the boat.

$boatX_0' \leftarrow cont(waterX) + cont(motorX);$

$boatY_0' \leftarrow cont(waterY) + cont(motorY);$

$F(boatX, boatX') = boatX_0' + \int_0^t boatX' \ (Convar\ (time))dt$

$F(boatY, boatY') = boatY_0' + \int_0^t boatY' \ (Convar\ (time))dt$

In this scenario, the sub-module boat is described with the hybrid language:
A::=x:=motorX, y:=motorY, X:= waterX, Y:= waterY, x1= boatX, y1= boatY
EQ::= F(boatX, boatX')=0|| F(boatY, boatY')=0
P::=P1||P2
P1::=EQ **until** g;
g::=s:= ((|Convar (boatX) – landingX |<= 0.1) and ((|Convar (boatY) – landingY| <= 0.1));
P2::=pause
G::=g&P2

The Water

There exits two kinds of situations of : whether considering the drift of water or not. Assume that the water is still, that is to say, we don't need to consider the velocity of water. Under this circumstance, the movement of the boat is very simple in that it is only determined by the motor. Otherwise, we have to think about both the water and the motor together. The input variables include the position of the boat, and the output variable is the water drift.

Firstly, let's consider the simplest situation:

Water1:
A::=x:=boatX, y:=boatY, X:= waterX, Y:= waterY
EQ::= X = 0,Y=0
P::=EQ **until** g;
g::=s:=absence(ε)

The second situation, that there exists a constant drift in a predifined direction.

Water2:
A::=x:=boatX, y:=boatY, X:= waterX, Y:= waterY
EQ::= X= 1, Y=-2.0
P::=EQ **until** g;
g::=s:=absence(ε)

In the third situation of the water's movement, there is a non-constant drift of the water that is defined by using sinus and cosine functions.

Water3:
A::=x:=boatX, y:=boatY, X:= waterX, Y:= waterY
EQ::=X <– sin(Convar (time) + 2.0), Y<– 1.5_cos(Convar (time)_4.0);
P::=EQ until g;
g::=s:=absence(ε)

The Boat Controller / Motor of the Boat

The boat is driven by the motor. It provides the boat with a constant velocity and we suppose that the velocity changes in real time. The coordinate of the boat is detected by the controller periodically. The next velocity is determined in accordance with the vector v from the boat position to the landing space. The direction of the vector is up to v, and we suppose that the velocity is constant every time at the length of 5.0. The input variables include the position of the boat, and the output variable is the next velocity of the boat water drift.

Contoller
v = 5.0;
timeOut = 2.0;
time=0.0;
A::=x:=boatX, y:=boatY, X:= motorX, Y:= motorY , x1:=landingX, y1:= landingY, x2:=diffX, y2:=diffY, l:=length
EQ::=x2=x1-x, y2=y1-y, l= exp(exp(x2,2) + exp(y2,2),0.5),
next(X)=(x2/l)*v,
next(Y)=(y2/l)*v,
time <–1.0;
P::=EQ until g;
g::=s:=Conver(time)>=timeOut

In this example, the three sub-components are driven respectively in a parallel way. So the overall system is exhibited as follows:

P::=Boat||Water||Controller

4 Conclusion

A Cyber physical system(CPS) combines computation component and physical process together, usually, also coupled with complicated clock interval and perplexed communication mechanism. In a cyber physical system, both the continuous and discrete dynamics are included. A system like this is called a hybrid system. According to the paper that we have referred to, It is obvious that the system is very difficult to be modeled. Jifeng He's theory is a breakthrough in this area in that it is a perfect programming language to model physical world and its interaction with the control program.

In this paper, I adapted this new language to the specification of hybrid systems. In my case study, a typical model concerning environment-controller interaction is built in which a motor boat and its sailing environment are included. Case study illustrates that using hybrid relation calculus to specify cyber physical systems can make every subsystem less complicated and more detailed, so that we can get easy access to observe the physical mechanism of the system. Hybrid relation calculus programming language is proved to be efficient and accurate.

Acknowledgements. This work is supported by the national natural science foundation of China under grant (No.61370082, No.61173046, No.91318301, No.91118008), natural science foundation of Guangdong province under grant (No.S2011010004905). This work is also supported by Shanghai Knowledge Service Platform Project (No.ZF1213).

References

1. COMPUTING E. Cyber-physical systems (2009)
2. Bingqing, X., Jifeng, H., Lichen, Z.: Specification of Cyber Physical Systems Based on Clock Theory (2013)
3. Simpson, H.R.: Four-slot fully asynchronous communication mechanism. Computers and Digital Techniques, IEE Proceedings E 137(1), 17–30 (1990)
4. He, J.: Hybrid Relation Calculus. In: 2013 18th International Conference on Engineering of Complex Computer Systems (ICECCS), p. 2. IEEE (2013)
5. Sanchez, C., Julian, B., Belleville, P., et al.: Applications of hybrid organic–inorganic nanocomposites. Journal of Materials Chemistry 15(35-36), 3559–3592 (2005)
6. Henzinger, T.A., Ho, P.H., Wong-Toi, H.: HyTech: A model checker for hybrid systems. In: Grumberg, O. (ed.) CAV 1997. LNCS, vol. 1254, pp. 460–463. Springer, Heidelberg (1997)
7. Dolbeau, R., Bihan, S., Bodin, F.: HMPP: A hybrid multi-core parallel programming environment. In: Workshop on General Purpose Processing on Graphics Processing Units (GPGPU 2007) (2007)
8. Blackburn, P., Seligman, J.: Hybrid languages. Journal of Logic, Language and Information 4(3), 251–272 (1995)
9. Bauer, K.: A New Modelling Language for Cyber-physical Systems. Verlag Dr. Hut (2012)
10. Zhu, L., Zhao, Y., Zhu, H., Xu, Q.: Towards a modeling language for cyber-physical systems. In: Liu, Z., Woodcock, J., Zhu, H. (eds.) Theories of Programming and Formal Methods. LNCS, vol. 8051, pp. 394–411. Springer, Heidelberg (2013)
11. Hennessy, M., Milner, R.: Algebraic laws for nondeterminism and concurrency. Journal of the ACM (JACM) 32(1), 137–161 (1985)
12. Man, K.L., Schiffelers, R.R.H.: Formal specification and analysis of hybrid systems. Dissertation Abstracts International 68(01) (2006)
13. Bingqing, X., et al.: Specification of Cyber Physical Systems by Clock. In: AST 2013, Yeosu, South Korea. ASTL, vol. 20, pp. 111–114 (2013)

14. Alur, R., Grosu, R., Hur, Y., Kumar, V., Lee, I.: Modular specification of hybrid systems in CHARON. In: Lynch, N.A., Krogh, B.H. (eds.) HSCC 2000. LNCS, vol. 1790, pp. 6–19. Springer, Heidelberg (2000)
15. Jifeng, H.: A clock-based framework for construction of hybrid systems. In: Liu, Z., Woodcock, J., Zhu, H. (eds.) ICTAC 2013. LNCS, vol. 8049, pp. 22–41. Springer, Heidelberg (2013)

A Network Delay Jitter Smoothing Algorithm in Cyber-Physical Systems

Gang He, XiangLi Zhang, HongMei Zhang, and YanYan Dai

GuiLin University of Electronic Technology, Guilin, 541004, China
112021207@mails.guet.edu.cn,
{xlzhang,hmzhang}@guet.edu.cn,
799582091@qq.com

Abstract. Network delay jitter variation in Cyber-Physical Systems (CPS) is one of the main factors which cause performance degradation or even system instability, how to smooth the network delay jitter or reduce the uncertainty is one of several major problems that face CPS. We have analyzed network delay jitter research status. A network jitter smoothing algorithm based on exponential smoothing prediction (ESP algorithm) was proposed in this paper. Experimental results show that the algorithm can predict network delay jitter size variation well, effectively reduce the changing network jitter outliers and smooth the network delay jitter. Compared with other algorithms, ESP algorithm has some advantages, and with no need for clock synchronization.

Keywords: CPS, Network delay jitter, jitter smoothing algorithm.

1 Introduction

Cyber-Physical Systems (CPS) are an organic integration of control technology, computing technology and communication technology. CPS will affect which humans interact with the physical world [1]. A lot of research has been done in this field, such as CPS architecture, controllers, actuators and sensors of the three elements of CPS [2-3]. Controller could send control commands to the actuators through the network, actuators execute the command at the right time. Network delay and jitter are the important factors that affect this process. The data buffer can absorb of irregular network jitter well. Reasonable data buffer size can smooth the network jitter and improve system real-time performance, high efficiency and stability .It has great significance. Data packets contain the control commands and other information. Actuators execute control command at the right time, change the state of physical objects. Buffer size is too large, data packets at the receiving end will wait for a long time and the system won't be real-time performance. If actuators miss the best execution time, the system will be damaged. Buffer size is too small, after packets are received, the command is executed quickly .That improve the system real-time performance, but the network jitter is too large so that lots of packets are lost .If the actuators can't receive the correct command, the system will be also damaged.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_8, © Springer-Verlag Berlin Heidelberg 2014

How to set the buffer size reasonably in the CPS has become the research focus at home and abroad. Researchers put forward a lot of kinds of network jitter smoothing and buffer size set algorithm.

In [4] the authors assumed that the upper limit of the size of the network latency is known, proposed an output feedback mechanism based on the maximum network delay and GPS clock synchronization.

In [5] the authors researched audio data packets delay jitter, analyzed the four kinds of adaptive adjustment algorithm, used autoregressive to predict the means and variances of network delay, then calculate the buffer size. The authors proposed algorithm that smoothed network jitter effectively. But the related parameters of the algorithm are under the network environment of particular experiments and the applicable scope of the algorithm is limited.

In [6] the authors put forward an integrated buffer, sensor sampling and the feedback control of end-to-end algorithm (PBD algorithm), and the algorithm according to the observed value of the network jitter to predict network delay. In their algorithm, every data packet has two control signals which would be called the regular control signal and the contingency control signal. This algorithm made up for the deficiency of simple feedback scheme: when accept time of packets is more than the set value, actuators can use contingency control signal control the plant. But the fluctuation of predicted value was very large and the performance of network jitter smoothing was not good.

According to the characteristics of CPS, in order to improve related algorithms, this paper introduced an exponential smoothing method to ESP algorithm which can approximate to predict the network delay jitter. This is the innovation of this paper. The experimental results show that the performance of the new proposed algorithm is better than MA algorithm's. The values calculated by the new algorithm follows the size of the observed values trend well. The accuracy of the predicted value can be strengthened. This paper is organized as follows. Sec.2 reviews the background and related work. Sec.3 introduces the exponential smoothing method (ESP algorithm) that we proposed. Sec.4 gives the experimental results and Sec.5 concludes the paper.

2 Background and Related Work

In [7] the authors proposed a scheme that calculated predicted values by observed values. The basic idea of their scheme is shown in figure 1. We can get the following equation:

$$\hat{f}_{pc}(i) = f_{pc}(i) - f_{pc}(i-1) = (t_j - t_{j-1}) - (t_i - t_{i-1}) \tag{1}$$

Where $f_{pc}(i)$ represents the transmission time of the i-th packet traverses the network from the physical system to the controller. t_i is the time of the i-th packet arrival physical system. t_j is the time of the i-th packet arrival controller. $\hat{f}_{pc}(i)$ is

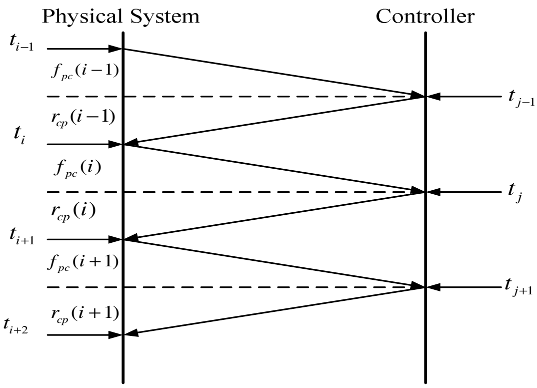


Fig. 1. Packet switching in network control system

the variation of $f_{pc}(i)$ .Equation (1) shows that $\hat{f}_{pc}(i)$ can be accurately calculated, and no need synchronous clock. The same can be concluded that:

$$\hat{r}_{cp}(i) = r_{cp}(i) - r_{cp}(i-1) = (t_{i+1} - t_i) - (t_j - t_{j-1}) \tag{2}$$

Where $r_{cp}(i)$ represents the transmission time of the i-th packet traverses the network from the controller to the physical system. $\hat{r}_{cp}(i)$ is the variation of $r_{cp}(i)$. Equation of ideal buffer size τ_i of the *i*-th packet as follows:

$$\tau_i = (t_i - t_{i-1}) + \hat{f}_{pc}(i) + \hat{r}_{cp}(i) \tag{3}$$

This method is called one way delay ($o\hat{w}d$). Vincenzo Liberatore et al. proved that the buffer size τ_i should be as close as possible to and greater than a packet round-trip time (RTT).we can get the following:

$$\tilde{\tau}_i = (t_i - t_{i-1}) + \hat{f}_{pc}(i) + \tilde{r}_{cp}(i)\ ,\ \tilde{r}_{cp}(i) \geq \hat{r}_{cp}(i) \tag{4}$$

Where $\tilde{r}_{cp}(i)$ is the prediction of $\hat{r}_{cp}(i)$. If $\tilde{r}_{cp}(i)$ can be predicted accurately, buffer size τ_i can be calculated and set reasonably.

3 Exponential Smoothing Prediction Method

The MA algorithm provided a closer approximation to the RTT and reduced frequent outliers. But the MA algorithm has a shortcoming: when delay jitter becomes very large, predicted values will become very big and not close the size of the network delay jitter any more. The MA algorithm will unable to reflect the real-time and superiority in this situation. In order to solve this problem, we introduced an exponential smoothing prediction method in our algorithm.

Exponential smoothing prediction method is a special kind of weighted method proposed by Brown (Robert G.. Brown). The basic mathematical model is shown as following, time sequence: $x_1, x_2, ..., x_n$, where n is the total number of installments for installments for the time sequence, the equation for the single exponential smoothing:

$$S_i^{(1)} = \alpha x_i + (1-\alpha) S_{i-1}^{(1)}, \hat{Y}_{i+1} = S_i^{(1)} \tag{5}$$

Where $S_i^{(1)}$ is the ith smooth value, α represents smoothing coefficient, $\alpha \in [0,1]$. $\hat{Y}_{i+1}$ is the (i+1)th predicted value. According to the Eq.5, we can get the following equation:

$$\hat{Y}_{i+1} = \alpha x_i + (1-\alpha) S_{i-1}^{(1)} = \alpha x_i + \alpha(1-\alpha) x_{i-1} + \alpha(1-\alpha)^2 x_{i-2} + ... + \alpha(1-\alpha)^{i-1} x_1 \tag{6}$$

Above equations show that the closer to the current time, the bigger the weights of observed value and the greater the effect on the predicted value, the further away from the current time, the smaller the weight of the observed value and the smaller the effect on the predicted value.

Exponential smoothing method considered the effect of all observed values in observation period. This effect decreases gradually according to time from near to far. Exponential smoothing method can make up for the inadequacy of MA algorithm.

4 Simulations and Results

In order to verify the validity of the buffer size setting, we carried out the experiment. We define a sliding window of size 40 in ESP algorithm. In the sliding window, we have 40 observed values $\hat{r}_{cp}(i)$. Each time we get a new $\hat{r}_{cp}(i)$, the sliding window slides backwards one value and update all observed values in the sliding window. We compare the $\tilde{r}_{cp}(i)$ setting obtained on ESP algorithm with EA algorithm and the packet RTT predicted PBD algorithm. We set α to 0.9. The average of observed value is 55 ms, observed values range from 45 ms to 65 ms, observed values space size is 400, most of values produced by random number except outliers. In order to show the difference of processing capacity of outlier between ESP algorithm and others (EA and PBD algorithm), we set many outliers: $\hat{r}_{cp}(30) = 70ms$, $\hat{r}_{cp}(30) = 70ms$, $\hat{r}_{cp}(70) = 80ms$, $\hat{r}_{cp}(90) = 75ms$, $\hat{r}_{cp}(150) = 25ms$, $\hat{r}_{cp}(200) = 125ms$.

We can see three experimental results in figure 2. The red curve consists of the observed values. The blue curve is consists of the predicted value. The black curve consists of that the predicted values deviate from the observed values. If black curve is less than zero, it indicates that the predicted value is less than the observed value which considered packet loss. So the black curve should be greater than zero and try to greater than zero. The PBD algorithm is to predict the packets' RTT, our algorithm and MA algorithm are to get the one way delay variation of the reverse path from the controller to the physical system ($\hat{r}_{cp}(i)$). Although they are differences, but they are calculated predicted values by observed values about network delay and jitter. We set

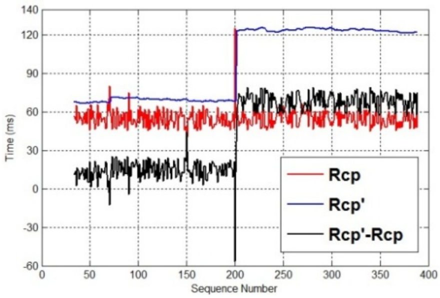


Fig. 2. Experimental results of MA algorithm

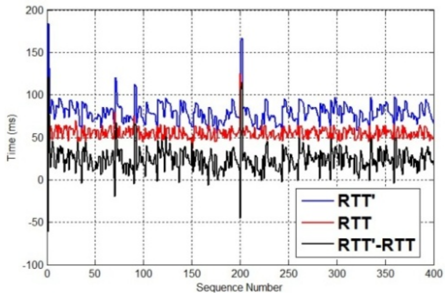


Fig. 3. Experimental results of PBD algorithm

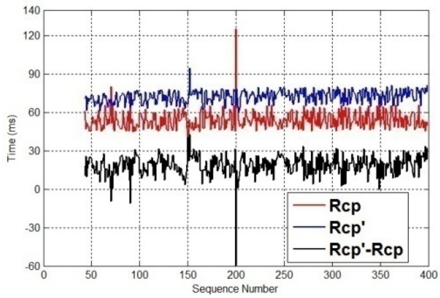


Fig. 4. Experimental results of ESP algorithm

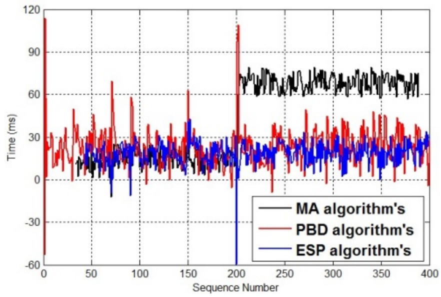


Fig. 5. The predicted values deviate from the observed values of three algorithms

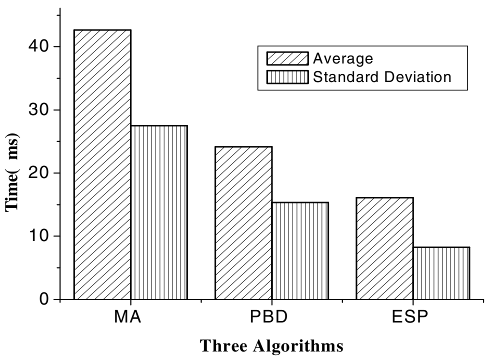


Fig. 6. The mean and standard deviation of the black curve

the outlier to $\hat{r}_{cp}(200) = 125ms$. The predicted values became very big and stay away from the observed values after 200th time sequence in Fig.2. The predicted values are larger than the others two results, the number of packet loss is more in Fig.3. Fig.4 shows that Exponential smoothing prediction method did not appear previous problem. In order to figure out the strength and weakness of the proposed scheme, we provided comparison results to related work, as shown in the Fig.5. The predicted values which calculated by MA algorithm deviate from the observed values is shown in black. The predicted values which calculated by PBD algorithm deviate from the observed values is shown in red. The predicted values which calculated by proposed scheme in this paper deviate from the observed values is shown in blue. Because the black curve contained information about the advantages and disadvantages of three algorithms, we calculated their mean and standard deviation, as shown in Fig.6.

5 Conclusion

This paper studied the existing network delay jitter smoothing algorithm in this paper, proposed ESP mechanism based on economic data predicted method. The experimental results show that, ESP mechanism can predict the network delay jitter variation better than the other two algorithms (PBD and MA), reduce the network jitter outliers effectively.

Acknowledgment. The authors would like to thank the Foundation items (NFS of China: No.61363031), GuangXiSci.&Tech. Development Foundation: No.GKG12118017-2C,GuilinSci. & Tech, Development Foundation: No.20120104-13, Innovation Project of Guangxi Graduate Education: YCSZ2012068).

References

1. Li, Z., Zhang, T., Zhang, Q.: Survery on the research of Cyber-Physcial systems. CPS 39(9), 25–31 (2011)
2. Lee, E.A.: Cyber physical systems: Design challenges. In: 2008 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC), pp. 363–369. IEEE (2008)
3. Rajkumar, R.R., Lee, I., Sha, L., et al.: Cyber-physical systems: the next computing revolution. In: Proceedings of the 47th Design Automation Conference, pp. 731–736. ACM (2010)
4. Seuret, A., Termens-Ballester, M., Toguyeni, A., et al.: Implementation of an internet-controlled system under variable delays. In: IEEE Conference on Emerging Technologies and Factory Automation (ETFA 2006), pp. 675–680. IEEE (2006)
5. Ramjee, R., Kurose, J., Towsley, D., et al.: Adaptive playout mechanisms for packetized audio applications in wide-area networks. In: 13th Proceedings IEEE Networking for Global Communications (INFOCOM 1994), pp. 680–688. IEEE (1994)
6. Liberatore, V.: Integrated Play-Back, Sensing, and Networked Control. In: INFOCOM (2006)
7. Al-Omari, H., Wolff, F., Papachristou, C., McIntyre, D.: An Improved Algorithm to Smooth Delay Jitter in Cyber-Physical Systems. In: Proceedings of the 8th International Conference on Embedded Computing and International Conference on Scalable Computing and Communications, pp. 81–86. IEEE, Dalian (2009)

A Novel Activity Recognition Approach Based on Mobile Phone

Lingxiang Zheng[1], Yanfu Cai[1], Zhanjian Lin[1], Weiwei Tang[1], Huiru Zheng[2], Haibin Shi[1,*], Bruce Liao[3], and Jolly Wang[3]

[1] School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
{lxzheng,shihaibin}@xmu.edu.cn
[2] School of Computing and Mathematics, University of Ulster, Jordanstown, BT370QB, U.K.
h.zheng@ulster.ac.uk
[3] Intel (China) Co., Ltd., Shanghai 200241, China
{bruce.liao,jolly.wang}@intel.com

Abstract. This paper presents a novel method for high-accuracy human activity recognition based on mobile phone acceleration sensors. Our approach includes two phases: the feature extraction phase and the classification phase. In feature extraction phase, we process tri-axial acceleration sensor data by combining the Independent Components Analysis (ICA) with the wavelet transform algorithm to get the features. In the classification phase, we apply the Support Vector Machine (SVM) algorithm to distinguish four types of activities (sitting, standing, walking and running). Experimental results show that the approach achieves an average accuracy of 98.78% over four types of activities, which outperforms the traditional method. The high accuracy indicates that this approach may facilitate the mobile phone based human activity recognition application.

Keywords: acceleration, Independent Components Analysis, wavelet transform, SVM, activity recognition.

1 Introduction

Human activity recognition (HAR) is an important topic for the development of robust machine learning techniques in pervasive computing applications. The recognition of human activities has been widely applied within various fields, particularly for medicare and health monitoring application.

One of the most popular HAR systems is the system which based on sensors especially the sensors embedded in the mobile phone. In such a system, the sensor data is very noise. How to remove the noise and extract the features properly is a crucial question in the activity recognition. Existing activity recognition systems based on accelerometer data apply statistical feature extraction, especially, time- or frequency-domain features. Mean, standard deviation, variance, correlation between axes, entropy, and kurtosis [1,2,3] , these time-domain features are widely used in traditional recognition process. The most popular frequency-domain features are

[*] Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_9, © Springer-Verlag Berlin Heidelberg 2014

high-pass filter(HPF) and mean filter(MF)[4] and Discrete Cosine Transform (DCT) [5]. Auto-regressive model coefficients which have been adopted have also yielded promising results [6].

Due to the noise in the data, these methods can hardly get high recognition accuracy. In this paper, we present a novel method to eliminate the noise from the data and try to get higher activity recognition accuracy. In our approach, we try to get rid of the impact of noise when extracting features by combining use of the Independent Components Analysis (ICA) algorithm and the wavelet transform algorithm. Considering the excellent performance of the Support Vector Machine (SVM) algorithm, we use the SVM as the classifier for activity recognition.

The rest of the paper is structured as follows: In Section 2, we discuss the detailed process of the proposed recognition strategy, including feature extraction and classification. Section 3 gives the experiment design to validate the effectiveness of the proposed method and experiment result. Finally, conclusions are given in Section 4.

2 Method

The framework of our activity recognition system is shown as Fig 1. With the combination of ICA and wavelet transform, the two most discriminating features can be extracted. The two extracted features and the three-axis acceleration value make up the feature vectors. After that we use mutil-class support vector machines [7,8] to distinguish the different activities of human.

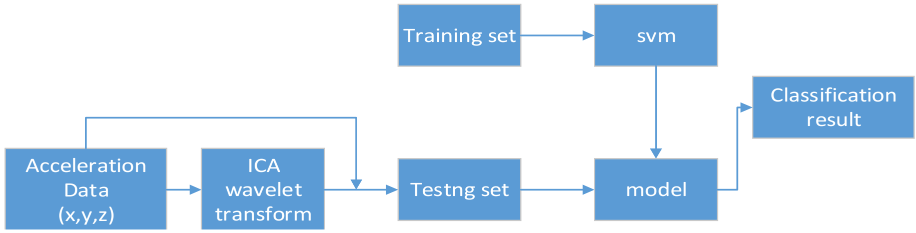


Fig. 1. The framework of our system

2.1 Independent Components Analysis

Independent component analysis (ICA) is a statistical and computational technique. ICA is superficially related to principal component analysis (PCA) and factor analysis. ICA is a much more powerful technique, however, capable of finding the underlying factors or sources when these classic methods fail completely.

In Independent Components Analysis, one assumes the m×p X matrix can be decomposed as

$$\mathbf{X} = \mathbf{AS}$$

Where **S** is the k×p matrix of k independent source signals (in rows), called the Independent Components, and **A** is the m×k mixing matrix. ICA aims at determining both **A** and **S**, knowing only **X**, by assuming that the source signals are mutually statistically independent, and that their mixing (yielding **X**) is linear. ICA, which aims to maximize independence between the extracted components, assumes that the “pure source signals” are less Gaussian than their mixtures (Central Limit Theorem) and so maximises the non-Gaussianity of the extracted ICs.

In this study, the fastICA [9] algorithm is applied for ICA calculations.

2.2 Wavelet Transform

Wavelet analysis is a relatively new signal processing tool and is applied recently by many researchers due to its strong capability of time and frequency domain analysis.

The definition of continuous wavelet transform (CWT) for a given signal x(t) with respect to mother wavelet $\Psi(t)$ is

$$CWT(\mathrm{a,b}) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

Where a is the scale factor and b is the translation factor. The Discrete wavelet transform (DWT) can be written as

$$DWT(\mathrm{m,n}) = \frac{1}{\sqrt[m]{a}} \sum_{k} x(k)\psi\left(\frac{k-nb_0a_0^m}{a_0^m}\right) \tag{2}$$

Where original a and b parameters (1) are changed to be the functions of integers m,n,k which is an integer variable and it refers to a sample number in an input signal. This technique can be used effectively for realizing non-stationary signals comprising of high and low frequency components, through the use of a variable window length of a signal. The ability of the wavelet transform to focus on short time intervals for high frequency components and long-time intervals for low frequency components improves the analysis of transient signals. For this reason, wavelet decomposition is ideal for studying transient signals and obtaining better current characterization and a more reliable discrimination. The wavelet transform technique is first applied, in order to, decompose the different current signals into a series of wavelet components, each of which is a time domain signal that covers a specific frequency band[10].

3 Experimental Results and Discussion

3.1 Experiment Design

Our experimental setup employs a Samsung Galaxy SIII cellphone running an Android 2.3 application to collect the data from accelerometer. The sample rate was set at 100Hz. The data from the accelerometer had the following attributes: time, acceleration along x axis, acceleration along y axis and acceleration along z axis. We collected data for four types of activities (sitting, standing, walking and running) over different days, as shown in Fig 2. We compare the traditional method which based on high-pass filter(HPF) and mean filter(MF).

The cellphone was naturally put in a pocket, and the sensors were not fixed with the body. So the data generated could conform to the real situation. Five subjects (two females and three males) were asked to perform these activities in two minutes for training data set and to carry cellphone in their daily life over several days for testing data set.

Fig. 2. The raw data of four activities (m/s^2)

Features were extracted from raw accelerometer data with ICA and wavelet transform. Here, the fastICA was adopted to process signal. The input of the fastICA were the x-, y-, z-axis acceleration components (x, y, z) and resultant acceleration. The output of the fastICA was set as four-dimensional output. We used a part of these data as the training dataset to train the SVM classifier. The test results show that the classifier reaches the best performance when the fourth dimension of the ICA output is used as the feature.

In a similar way, the fastICA was reset as one-dimensional output. And then we also used wavelet transform to process the output value of fastICA for de-noising. After that it would be the second feature. The two features and the x-, y-, z-axis acceleration components (x, y, z) made up of the feature vector for classification system.

In our case, we applied the WEKA toolkit, a collection of machine learning algorithms for activity recognition tasks [11]. The SVM was used as the classifier for activity recognition and the default parameters were applied. Then the leave-one-subject-out validation strategy was adopted to evaluate the classifiers' performance. One subject was regarded as the testing data set for classifiers, except which, all other data trained the classifiers. This process was repeated for all subjects.

3.2 Results

This section describes experiment results with the proposed activity recognition system. Table 1 shows that, the recognition performance of SVM algorithm with the proposed features vector is excellent (average accuracy 98.78%), in leave-one-subject-out test. All activities are close to 100% except the standing (96.6%).

Table 1. The accuracy of activity recognition

Method	Sitting	Standing	Walking	Running	Average
ICA+Wavelet	100%	96.6%	99.2%	99.3%	98.78%
HPF+MF	100%	81.6%	99.1%	97.7%	94.60%

The performance of activity recognition is evaluated from three indicators-recall rate, precision and F-measure.

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search.

$$precision = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|} \tag{3}$$

As is shown in Fig 3 (a), the precision of standing activity using wavelet and ICA is 15.0 % higher than the same activity using high-pass and mean filter.

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$recall = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|} \tag{4}$$

As is shown in the Fig 3 (b), the recall rate of two methods are close. But the difference lies between the sitting and standing: the former is approximately 1.5% higher than the latter.

In statistics, the F-measure is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.

$$F = 2 \bullet \frac{precision \bullet recall}{precision + recall} \tag{5}$$

In Fig 3 (c), we can find out that the F-measure of wavelet and ICA is 8.08% higher than the high-pass and mean filter.

In summary, ICA and wavelet transform method is superior to high-pass filter and mean filtering method in terms of recall, precision and F-measure.

In order to find out which activities are relatively harder to be recognized, we analyze the confusion matrices, which give information about the actual and predicted classification results given by the classifiers. Table 2 shows the aggregate confusion matrix for the SVM classifier based on all eleven trials of leave-one-subject-out validation. It can be seen that walking is often confused with running and is in general hard to recognize.

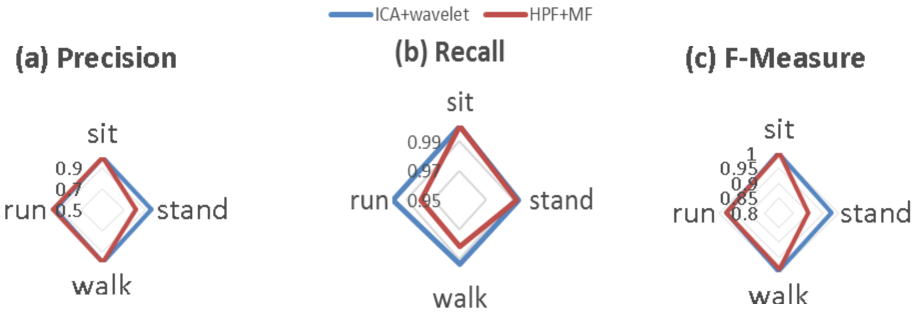


Fig. 3. The evaluation of classifier

Table 2. Confusion matrices of activity recognition

	Sitting		Standing		Walking		Running	
Method	ICA+ Wavelet	HPF+ MF	ICA+ Wavelet	HPF+ MF	ICA+ Wavelet	HPF+ MF	ICA+ Wavelet	HPF+ MF
Sitting	115998	115998	64	60	15	25		22
Standing	0	0	2480	2095	3	7	0	0
Walking	0	0	24	413	29376	29345	22	49
Running	0	0	0	0	218	235	3047	2998

4 Conclusion

Due to the amount of noise in the data, it is a hard work to achieve the activity recognition using the acceleration sensors in the mobile phone. In this paper, we presented a novel two-phase-method to do the activity recognition. In the first phase, we did the feature extraction by eliminating the noise using an algorithm combing the ICA algorithm with the wavelet transform algorithm. In the second phase, we used the SVM to classify the activities based on the feature extracted from the first phase. In comparison with the method using filter for features extraction, our approach is more efficient for data pre-processing. This efficient feature vector improves the performance of SVM for activity recognition. The experiment result demonstrated that the accuracy of activity recognition for four daily activities can reach up to 98.78%. It indicates that the combination of the ICA algorithm with wavelet transform algorithm is more effective in feature extraction which can further improve the performance of activity recognition.

References

1. Lara, Ó.D., Pérez, A.J., Labrador, M.A., Posada, J.D.: Centinela: A human activity recognition system based on acceleration and vital sign data. Pervasive and Mobile Computing (2011)

2. Tapia, E.M., Intille, S.S., Haskell, W., Larson, K., Wright, J., King, A., Friedman, R.: Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor. In: Proc. Int. Symp. on Wearable Comp. Citeseer (2007)
3. Kao, T., Lin, C., Wang, J.: Development of a portable activity detector for daily activity recognition. In: IEEE International Symposium on Industrial Electronics (ISIE 2009), pp. 115–120. IEEE (2009)
4. Chen, Y., Yang, J., Liou, S., Lee, G., Wang, J.: Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Appl. Math. Comput. 205(2), 849–860 (2008)
5. He, Z., Jin, L.: Activity recognition from acceleration data based on discrete consine transform and SVM. In: IEEE International Conference on Systems, Man and Cybernetics (SMC 2009), pp. 5041–5044. IEEE (2009)
6. He, Z., Jin, L.: Activity recognition from acceleration data using AR model representation and SVM. In: 2008 International Conference on Machine Learning and Cybernetics, pp. 2245–2250. IEEE (2008)
7. Buhmann, J.M., Busetto, A., Vezhnevets, A.: Statistical Learning Theory (2013)
8. Hastie, T., Tibshirani, R., Friedman, J.J.H.: The elements of statistical learning, vol. 1. Springer, New York (2001)
9. Koldovsky, Z., Tichavsky, P., Oja, E.: Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound. IEEE Transactions on Neural Networks 17(5), 1265–1277 (2006)
10. Malathi, V., Marimuthu, N.S.: Wavelet transform and support vector machine approach for fault location in power transmission line. International Journal of Electrical and Electronics Engineering 4, 272–277 (2010)
11. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)

Indoor Pedestrian Navigation with Shoe-Mounted Inertial Sensors

Xianchao Zheng[1], Hui Yang[1], Weiwei Tang[1], Shuxiang Pu[1], Lingxiang Zheng[1,*], Huiru Zheng[2], Bruce Liao[3], and Jolly Wang[3]

[1] School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
lxzheng@xmu.edu.cn
[2] School of Computing and Mathematics, University of Ulster, Jordanstown, BT370QB, U.K.
h.zheng@ulster.ac.uk
[3] Intel (China) Co., Ltd., Shanghai 200241, China
{bruce.liao,jolly.wang}@intel.com

Abstract. High accuracy in indoor navigation with shoe-mounted inertial sensors attracts a lot of researches in the last decades. In this paper, we build an indoor navigation system using a kind of estimation architecture with shoe-mounted inertial sensors. The architecture consists of zero velocity update (ZUPT) and extended Kalman filter (EKF). The ZUPT during the rest phase of a pedestrian's foot can be used together with an EKF. The real time EKF runs to estimate the drift error and non-linear error growth of accelerometers and gyroscopes. The algorithm is inspected and verified on an experiment board. It presents a good performance. The position error of our algorithm is below 1% of the actual total traveled distance. It is feasible to obtain a long-term stability and high accuracy in an indoor navigation scenario.

Keywords: indoor pedestrian navigation, inertial sensors, zero velocity update, extended Kalman filter.

1 Introduction

Indoor pedestrian navigation became a hotspot of navigation system research over the recent decades. There are two main methods used in indoor pedestrian navigation: wireless networking technology [1][2][3] and inertial sensors based approaches. Although wireless networking technology is a simple method, its performance is poor. On the contrary, the inertial sensors based approaches show a better performance [4][5][6][7][8][9]. This paper mainly focuses on shoe-mounted inertial sensors based navigation system for pedestrians. In a shoe-mounted sensors based indoor pedestrian navigation system, the heading drift and nonlinear error growth become a serious problem. In the paper [4][5][6][7], particle filtering is employed to consider the nonlinear dynamics of the human by means of a dedicated pedestrian movement model. As proposed in [8][9], zero velocity update (ZUPT) allows the EKF to correct the velocity error after each stride, breaking the cubic-in-time error growth and replacing it with an error accumulation that is linear in the number of steps. In the paper [8][10], the

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_10, © Springer-Verlag Berlin Heidelberg 2014

compass and pedestrian dead-reckoning are used to reduce the heading drift. In order to guarantee the accuracy, additional sensors and priori information are used in the navigation systems. The paper [11] proposes a motion model based on maps and floor-plans to weigh the possible headings of the pedestrian as a function of the local environment. The paper [12] presents an expansion to odometer which is based on simultaneous localization and mapping (SLAM) for pedestrians. These shoe-mounted inertial sensors based approaches which combine ZUPT, EKF, PF or additional conditions get a high accuracy in indoor navigation.

However, these methods are too complex or have some drawbacks. The disadvantage of additional compass is that it will be disturbed by the metal in the building. And the priori information method, such as the indoor map, is not suitable for the indoor scenario which the information are not exist. Due to the high computational complexity, the particle filtering is hard to update the position in real time. Moreover, most of these approaches usually use high-end sensors for keep high accuracy, which hazard the use of these approaches in real world. Considering the situation mentioned above, we try to build a low-cost, high accuracy and real time indoor navigation system which is based on a foot-mounted 6DOF strapdown inertial platform. The architecture of the system includes a zero velocity update (ZUPT) component and a real time extended Kalman filter (EKF) component. We employ a multi-condition ZUPT to guarantee the accuracy and try to reduce the unbounded non-linear error growth of velocity and location. An extended Kalman filter (EKF) combines with Quaternion algebra algorithm are used to estimate and correct the problem of non-linear error growth in inertial integration over time.

The rest of this paper is organized as follows: Section 2 introduces details on method to integrate inertial sensors, including ZUPT and EKF. Section 3 presents the experiments and results. Section 4 gives a conclusion of the paper, finally.

2 Method

We employ an estimation architecture based on extended Kalman filter. As illustrated in Fig.1, the position and heading are calculated using inertial sensors. The data is processed by ZUPT to detect whether the foot is stationary on the floor. The ZUPT triggers the Kalman filter which is used to estimate and correct the non-linear error growth of the position and heading over time.

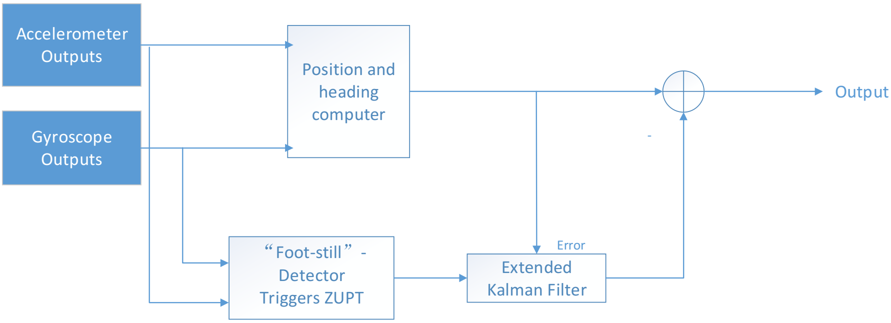


Fig. 1. System architecture

2.1 Zero Velocity Update

ZUPT is conducted as an effective method to control and eliminate errors. ZUPT is triggered when the foot is stationary on the ground. Our ZUPT algorithm is multi-condition by using information which output by accelerometers and gyroscopes, with an order filter to make the detection result accurate. Although the paper [10] shows the three conditions (C1, C2 and C3) to determine whether the foot is stationary on the floor, the judgment is inaccurate for the slight jitter with large noise of local heading caused by the gyroscope. In order to solve this problem, we use the following conditions:

$$C1 = \begin{cases} 1 & a_{\min} < |a_{k-total}| < a_{\max} \\ 0 & otherwise \end{cases} \tag{1}$$

$$C2 = \begin{cases} 1 & \sigma_{a_k}^2 > \sigma_{a-\min} \\ 0 & otherwise \end{cases} \tag{2}$$

$$C3 = \begin{cases} 1 & |w_{k-total}| < w_{\max} \\ 0 & otherwise \end{cases} \tag{3}$$

$$C4 = \begin{cases} 1 & \sigma_{w_k}^2 > \sigma_{w-\min} \\ 0 & otherwise \end{cases} \tag{4}$$

Where $a_{k-total}$ represents the magnitude of the acceleration at time k, $\sigma_{a_k}^2$ is the local acceleration variance, $w_{k-total}$ represent the magnitude of the heading, $\sigma_{w_k}^2$ is the local heading variance.

Only when the four logical conditions are satisfied simultaneously, ZUPT would be triggered. Thus, we employ a logical “AND” and the result is filtered out with a median filter.

2.2 Extended Kalman Filter

The task of EKF in this paper is to estimate the error of the state $\hat{x}_k$. It is triggered only when the foot is detected to be stationary on the ground. In other words, the “AND” result of four conditions in ZUPT would trigger the extended Kalman filter. We can get the estimation results after predicting and updating the state.

The state transition model x_k which we have used could only estimate the standard errors, is consist of position $(x_k\ y_k\ z_k)$, velocity $(v_{xk}\ v_{yk}\ v_{zk})$ and attitude matrix $(roll_k\ \ pith_k\ \ heading_k)$ of the pedestrian respectively:

$$\hat{x}_k = \left[x_k\ y_k\ z_k\ v_{xk}\ v_{yk}\ v_{zk}\ roll_k\ \ pith_k\ \ heading_k \right] \tag{5}$$

And the predicted state variable $\hat{x}_{k|k-1}$ and covariance of error state $P_{k|k-1}$ would be:

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + w_k \tag{6}$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{7}$$

$$F_k = \frac{df_k(x)}{dx}\Big|x_{k-1|k-1} \tag{8}$$

Here, Q_k is the covariance matrix of processing Gaussian white noise w_k . F_k and H_k are the local linearization of $f_k(\cdot)$ and $h_k(\cdot)$, which are computed as Eq.9 and Eq.10. The navigation state transition model $f_k(\cdot)$ is a non-linear function of the state variable x_k . The nonlinear observation transition function $h_k(\cdot)$ character the relationship between position and velocity.

$$F_k = \begin{bmatrix} 0 & I_{3\times3} & 0 \\ 0 & 0 & St \\ 0 & 0 & 0 \end{bmatrix} \tag{9}$$

$$H_k = \begin{bmatrix} 0 & I_{3\times3} & 0 \end{bmatrix} \tag{10}$$

where $I_{3\times3}$ is an identity matrix; St is a symmetric matrix with elements defined as the body-frame angular rates in Quaternion algebra, find detail in [13]:

$$St = \begin{bmatrix} 0 & -f(3) & f(2) \\ f(3) & 0 & -f(1) \\ -f(2) & f(1) & 0 \end{bmatrix} \tag{11}$$

The matrix f is coordinate transition matrix of acceleration.

$$\begin{cases} R_\varphi = \begin{bmatrix} \cos\phi\cos\lambda & \sin\phi\cos\lambda & -\sin\lambda \\ \cos\phi\sin\lambda\sin\omega - \sin\phi\cos\omega & \cos\phi\cos\omega + \sin\phi\sin\lambda\sin\omega & \cos\lambda\sin\omega \\ \sin\phi\sin\omega + \cos\phi\sin\lambda\cos\omega & \sin\phi\sin\lambda\cos\omega - \cos\phi\sin\omega & \cos\lambda\cos\omega \end{bmatrix} \\ f = R_\varphi \times acc \end{cases} \tag{12}$$

where R_φ is the transition function between body coordinate and earth coordinate, it is calculated in the quaternion algebra; acc is the data comes from accelerometer.

Then we would update the predicted state $\hat{\mathrm{x}}_{\mathrm{k|k}}$ and covariance $P_{k|k}$ with the optimal Kalman gain K_k and measurement residual $\tilde{y}_k$.

$$\hat{\mathrm{x}}_{\mathrm{k|k}} = \hat{\mathrm{x}}_{\mathrm{k|k-1}} + \mathrm{K}_k \tilde{y}_k \tag{13}$$

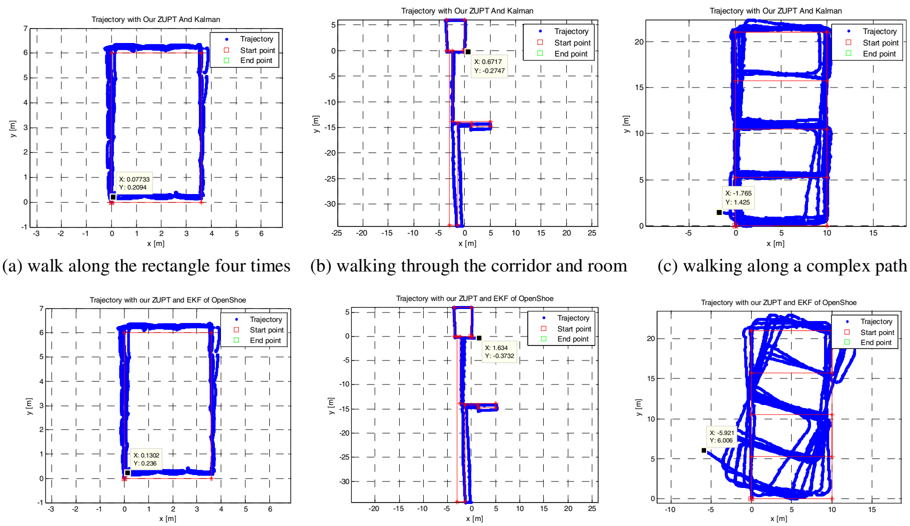
$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \tag{14}$$

3 Experiments and Results

We employ a simple and cheap full six degree of freedom (6DOF) inertial sensor (MPU6000) to verify the algorithm. The tester walks through a building only with the shoe-mounted inertial navigation system (INS). The INS data are recorded in 100Hz.

We make a comparison with the ZUPT and EKF proposed by OpenShoe [9] respectively. We test the ZUPT algorithm performance with two data sets: the OpenShoe dataset[1] and the dataset of ourselves[2]. Table 1 shows that both approaches can process the data set of OpenShoe and gets a similar result. However, the ZUPT of OpenShoe fails when process the dataset of us. Our ZUPT has a good performance on controlling and eliminating errors with high accuracy.

Fig. 2 shows the results of EKF algorithms used by OpenShoe and the one used in the paper. There are three trajectory of the pedestrian walking in our office building. The first trajectory is a walk along the rectangle(3.6m×6m) four times. The second one is the pedestrian walking through the corridor and room. The third is a trajectory of walk in a rectangle with a complex path for about 9 minutes. The red line indicates the real path of the trajectory. Fig2.(a)(b)(c) show the result of our algorithm, Fig2.(d)(e)(f) present the result of the OpenShoe's EKF algorithm with our ZUPT(OpenShoe's ZUPT algorithm fails to process the dataset). The detail information of the three trajectories is shown in Table2.The starting and ending point of all the trajectories is (0,0).



(a) walk along the rectangle four times (b) walking through the corridor and room (c) walking along a complex path

(d) walk along the rectangle four times (e) walking through the corridor and room (f) walking along a complex path

Fig. 2. Experimental result of Comparing EKF

1 `http://www.openshoe.org/`

2 `https://github.com/lxzheng/ShoeMountIMU-Dataset`

Table 1. Comparing Results of ZUPT

Trajectory		a	b	c	OpenShoe
Actual Steps		64	87	487	Unknow
Estimated Steps	Our ZUPT	64	87	487	43
	OpenShoe	22	23	3	41

Table 2. Comparing Results of EKF

Test		Actual Distance/m	Mean Stride Length/m	Estimated Mean Stride Length/m	End Location (x/m,y/m)	Estimated Distance/m	Drifting Error	Distance Error
Trajectory with Our EKF	a	76.8	1.2	1.1932	(0.07733,0.2094)	76.3619	0.57%	0.29%
	b	103.8	1.1931	1.1861	(0.6717,-0.2747)	103.1903	0.70%	0.59%
	c	575	1.1807	1.1737	(-1.765,1.425)	571.5676	0.39%	0.60%
Trajectory with The EKF of OpenShoe	d	76.8	1.2	1.1932	(0.1302,0.236)	76.3628	0.57%	0.35%
	e	103.8	1.1931	1.866	(1.634,-0.3732)	103.236	1.61%	0.54%
	f	575	1.1807	1.1719	(-5.921,6.006)	570.7393	1.47%	0.74%

As Table 2 shows, the average total traveled distance error and drifting error of our indoor navigation algorithm are lower than 1%. Comparing with the real path, we eliminate the nonlinear error growth and reduce heading drift successfully. And the unbounded error when the pitch angles are around $\pm 2/\pi$ is solved by the applying of Quaternion algebra, which uses four Euler parameters instead of three Euler angles. Comparing Fig2.(a)(b)(c) with Fig2.(d)(e)(f), the proposed EKF get a lower drifting error.

4 Conclusion

An estimation architecture based on shoe-mounted INS is presented in this paper. We employ a foot-mounted 6DOF inertial platform to detect a pedestrian's movement in an indoor scenario. According to the experiments and results, we successfully get the pedestrian's walking path without any source of absolute position information such as GNSS, additional sensors or heading in the dead-reckoning. The result shows that the integration algorithm has a good performance on trajectory matching with the EKF and ZUPT. The EKF and ZUPT could eliminate the nonlinear error growth and reduce the heading drift effectively. The Quaternion algebra combined in the EKF is able to calculate quickly for real time system. The algorithm can be used in indoor navigation which requires real time and low cost.

References

1. Bahl, P., Padmanabhan, V.N.: User location and tracking in an in-building radio network. Microsoft Research (February)
2. Ferris, B., Fox, D., Lawrence, N.D.: Wifi-slam using gaussian process latent variablemodels. In: IJCAI, vol. 7, pp. 2480–2485 (2007)
3. Bruno, L., Robertson, P.: Wislam: Improving footslam with wifi. In: 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–10. IEEE (2011)

4. Krach, B., Robertson, P.: Integration of foot-mounted inertial sensors into a Bayesian location estimation framework. In: 5th Workshop on Positioning, Navigation and Communication (WPNC 2008), pp. 55–61. IEEE (2008)
5. Robertson, P., Angermann, M., Krach, B.: Simultaneous localization and mapping for pedestrians using only foot-mounted inertial sensors. In: Proceedings of the 11th International Conference on Ubiquitous Computing, pp. 93–96. ACM (2009)
6. Angermann, M., Robertson, P.: Footslam: Pedestrian simultaneous localization and mapping without exteroceptive sensorshitchhiking on human perception and cognition. Proceedings of the IEEE 100(13), 1840–1848 (2012)
7. Angermann, M., Robertson, P., Kemptner, T., Khider, M.: A high precision reference data set for pedestrian navigation using foot-mounted inertial sensors. In: 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6. IEEE (2010)
8. Foxlin, E.: Pedestrian tracking with shoe-mounted inertial sensors. IEEE Computer Graphics and Applications 25(6), 38–46 (2005)
9. Nilsson, J.O., Skog, I., Handel, P., et al.: Foot-mounted INS for everybody-an open-source embedded implementation. In: 2012 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 140–145. IEEE (2012)
10. Jimenez, A.R., Seco, F., Prieto, J.C., Guevara, J.: Indoor pedestrian navigation using an ins/ekf framework for yaw drift reduction and a foot-mounted IMU. In: 2010 7th Workshop on Positioning Navigation and Communication (WPNC), pp. 135–143. IEEE (2010)
11. Kaiser, S., Khider, M., Robertson, P.: A human motion model based on maps for navigation systems. EURASIP Journal on Wireless Communications and Networking 2011(1), 1–14 (2011)
12. Robertson, P., Angermann, M., Khider, M.: Improving simultaneous localization and mapping for pedestrian navigation and automatic mapping of buildings by using online human-based feature labeling. In: 2010 IEEE/ION Position Location and Navigation Symposium (PLANS), pp. 365–374. IEEE (2010)
13. Hall, J.K., Knoebel, N.B., McLain, T.W.: Quaternion attitude estimation for miniature air vehicles using a multiplicative extended Kalman filter. In: 2008 IEEE/ION Position, Location and Navigation Symposium, pp. 1230–1237. IEEE (2008)

Presentation Interface Based on Gesture and Voice Recognition

Jinuk Kim, Sehoon Kim, Kwangjin Hong, David Jean, and Keechul Jung

School of Media, Soong-sil University, Seoul, South Korea
{jinuk,sehoon,hongmsz,jeandavid,kcjung}@ssu.ac.kr

Abstract. In this paper, we introduce a Kinect based interface that recognizes gestures and voice. We have developed an interface to control presentations such as speeches or lectures. It is possible to receive the coordinates of the body, and recognize gestures and positions of the hand. Data received by the camera in Kinect are used to create a hook between the user hand and a presentation application such as Microsoft Powerpoint. Our interface is able to recognize grip and push gestures from the presenter. The result of this gesture recognition generates a signal to the presentation application, such as shortcuts to change slides or make use of additional tools. It is also possible to start and end the presentation by voice using our voice recognition tool. Additionally we show some tools that not only change the slides, but also provide more options to the presenter such as memo tools to directly highlight some parts of a slide, and even an eraser. This paper describes all the methodology and presents the result of our tests session. We are effectively able to improve the presentation capability of the presenter and think that such interface can be commercialized for presentation and other type of use.

Keywords: Gesture recognition, Voice recognition, Presentation interface.

1 Introduction

After 1980, generalized GUI interfaces consist of restricted environments that only use mouse. But the development of input and output devices pushed interaction from a computer centric paradigm toward a more human and natural nature. MIT media lab started graphic interface development research with Chris Schmandt's voice and gesture in 1979[1]. This project has attracted more attention on gesture recognition as an alternative for eliminating the inconvenience of using input devices in specific contexts.

Unlike basic typing and mouse control, it becomes necessary to explore natural interaction between user and system for game and contents. In such areas, interactive systems consist of spaces and sensors that generate interactive and immersing experiences. It is for example used on visitors moving in a museum that could see some specific content that pop-up just by moving or using some gestures. We believe that it would create a better experience.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_11, © Springer-Verlag Berlin Heidelberg 2014

The development of interaction between computers and human can provide a lot of help in the everyday tasks, for example, in the case of presentations. When giving a presentation, natural behavior from the speaker is an important aspect that heightens the audience's concentration. When controlling a presentation using a device, these actions create an unnatural pause that can decrease the concentration of the audience.

This paper explores how to control a Powerpoint project by using Kinect sensor's skeleton position recognition in order to track hand position and recognize specific gestures. We use hand position to control the slides in Powerpoint while progressing the presentation naturally. We also developed a memo function that allows the user to write any content directly on the slide that is shown in Powerpoint. The user can start a presentation by using its voice and end it the same way. While doing the presentation, it is possible to move to another slide naturally. The presenter can improve its presentation capability by using a better interface environment.

2 Related Research

Especially in the fields related to computer vision, 2010 has seen a large number of studies based on movement and speech recognition, and still today, a lot are in progress. Kinect allows many researches to easily use a human skeleton, and such tool has generated many commercial applications.

Osunkoya[2] used the skeleton in order to obtain the coordinates of the hands and other specific parts (such as head or shoulders). The distances or respective positions of these different parts are then used to send events such has mouse move, or right and left clicks. These actions are then used to control Powerpoint. This method requires the users to place their hands at uncomfortable positions such as above their head, to insure the system can easily recognize the positions and send the right signals. Also, when moving from the position of one action to the position that recognizes another action, the risks to trigger another undesired action in between are high.

Gabacia[3] has developed a system that allows using simultaneously Kinect and Wiimote in order to control the mouse cursor during a presentation. This system allows the presenter to use voice recognition, Wiimote acceleration sensor and some buttons. Although this method uses Kinect and Wiimote, the control of the movement of the cursor is mainly based on the Wiimote device and doesn't make use of gesture recognition.

Another research proposed by Ren focuses on recognizing hand gestures using Kinect RGB camera to obtain the hands coordinates[4]. It is possible to recover the outline of the hand and make a timed series of the entire outline on each image to extract some feature vectors. This method shows 90% accuracy. However, the user must wear a black band on his hand to help the system reading the hand position correctly. Also, this method needs to compare all the raw data to all the features that are stored in a library, which is very computationally expensive.

Codeplex[5] is a system that uses the user body in order to control a slideshow. In Codeplex, the distance between the head and the guiding hand is constantly measured, and is used to trigger some functions. However, the lack of clear start and end of the gestures makes it difficult and causes many undesired actions. In order to fix this type of error, we believe that the beginning and end of the gestures must be clearly caught by the system.

Also, SmartTV proposes a personalized service that is based on gesture recognition. It is possible to control the latest generation of Samsung smart televisions with the new technology named Smart Interaction which actively make use of voice and hand gesture recognitions[6]. Recognition of the gestures uses the coordinate's values of the moving hands when they are in a grip posture. The main problems come as the system recognizes gestures that are not intended by the user, and process and undesired action.

Lately, lots of systems include more and more voice recognition in addition to gesture recognition, for researches as well as for commercial purposes. Apple's Siri[7], or Samsung's SVoice[8] offer services that make use of high level of technology in voice recognition.

3 System Structure and Functions

Kinect for Xbox 360 uses an input that consists of an RGB camera, an IR sensor, a microphone and a Tilt Motor. IR sensor can measure object's depth different from camera. Recognition range of Kinect is from 0.8m to 4m. Up and down movement of Kinect are recognized using tilt motor and microphone can recognize voice. Using Microsoft Kinect SDK v1.7 library it is possible to extract voice and skeleton information. It allows to recognizing gestures through skeleton information X, Y, Z coordinates. Microsoft Powerpoint slide show can be controlled by using mouse and keyboard. In this paper, we use Kinect to obtain user's body coordinates. Using the Kinect, we are able to recognize the position and gesture using coordinate values of a part of the body of the user. Especially, basic slide movement as well as memo, erase can be controlled.

Detected coordinates with the Kinect library are measured on the pixels of the Kinect common screen space. For this project, we have to manage 2 different spaces: one we call the application space and another we call the Kinect space. We first extract the coordinates of the hand from the Kinect skeleton space (K_x, K_y) and convert it to obtain the coordinates inside our application space (x, y). We are then able to process the movement.

However, these spaces have different coordinates systems and base units. The Kinect skeleton space has its base(0,0 coordinate) at the center of the Kinect view, in which the skeleton is displayed, and the unit used is meter. This space width and height are respectively called K_w and K_h. But the application space, has its base at the top left part of the screen, and the basic unit is pixel. Application space width and height are respectively called S_w and S_h. So we have to convert the points from one space to the other. The x and y coordinates that we need are extracted following the equation.

$$\begin{bmatrix} x \\ \mathrm{y} \end{bmatrix} = \begin{bmatrix} S_w/K_w & 0 & -S_w/2 \\ 0 & S_h/K_h & -S_h/2 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} K_x \\ K_y \\ 1 \end{bmatrix} \tag{1}$$

Kinect camera uses bilateral symmetry, like when a person looks at mirror. We set the Kinect in front of user and a screen is backside of the user. In this situation user moves the right hand and the cursor moves to left in the screen.

It means that the mouse controls is inverse to the speaker. To match a movement of the speaker in Kinect camera's coordinates, we extract the coordinates of the cursor ($C_{x,}$ C_y).We use equation2 to change the coordinates.

$$\begin{bmatrix} C_x \\ C_Y \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} * \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} S_w \\ 0 \end{bmatrix} \tag{2}$$

We extract the coordinates value of pixel unit and we can control the mouse cursor to follow the right hand's x, y coordinates. We also developed functions that recognize user special hand movements to go to previous slide or next slide. Editing function is divided into two parts, one is memo function and the other is eraser function.

Each function specified by the user's gestures is recognized at the same time internally, and the corresponding shortcut is then executed. In a slideshow situation the user can invoke functions such as going to next slide, previous slide, staring the pen tool or eraser tool, drag event and click event.

We designed specific gestures for each of the functions we developed. For example, the grip gesture followed by a sweep to the right allows the user to display the next slide of the presentation. The following Table 2 shows all the functions and their corresponding gestures

Table 1. Gesture for controlling function

Function	Gesture
Mouse Click	Grip gesture
Mouse Click Release	Release gesture
Next Slide	Grip & sweep right
Back Slide	Grip & sweep left
Mouse Cursor Change	Press gesture

Move to the next slide is performed when right hand is in gripping state and release grip after it moves to the right. The move to the right is recognized when the position of the hand on the following frame shows a significant difference in X coordinate to the current position. The same process is used for the sweep to the left, however, we use the left hand to recognize it, and the X position is the opposite than its right counterpart. For the push and grip gestures, we are doing a similar process but using the Z coordinate this time. Microsoft Kinect SDK does not recognize the grip posture of a hand at a satisfying rate. It often happens that a grip posture is recognized although it was not intended by the user. It is especially true when the hand overlap with the body in the camera image. This is why we developed a more robust algorithm that recognizes the grip posture only when the user keeps is hand closed for 10 consecutive frames.

Our program uses another thread to process the voice recognition. We implemented in a multitasking way so that updates are done in real time, and it is possible to use other functions during the execution of the voice recognition. Voice recognition and gesture recognition use same Microsoft Kinect SDK, but internally voice recognition uses Speech Recognition Language pack(en-US) of the Microsoft Speech Library[9]. The voice recognition system searches inside the Speech Recognition Language pack(en-US) engine for specific words that we registered, and check the similarities with the words coming from the user. Words that are registered during the operation are recognized using the voice input form the microphone integrated in Kinect.

Before saying “Start Presentation” Kinect doesn’t recognizes any movement. After voice is recognized, the slideshow and the gesture recognition start. On the contrary, when saying “End presentation” the slideshow and gesture recognition end. Gesture and voice recognitions can be efficient ways to control a presentation through user’s gesture by using Kinect. However, we have to make sure that the Powerpoint application is the application running at the foreground insuring the shortcuts generated are passed to this application. For this, our interface uses hooking to control the mouse from received coordinate value from Kinect.

In computer programming, the term hooking covers a range of techniques used to alter or augment the behavior of an operation system, of applications, or of other software components by intercepting function calls or messages or events passed between software components. Code that handles such intercepted function calls, events or messages is called a ‘hooking’[10].

Use hooking, Powerpoint can be controlled by Kinect that recognize gesture like basic window mouse event handler.

4 Experimental Results

Our system experiments have been done in a basic lecture room. Distance between the Kinect and the user is 3.5m in the experimental space, and the range of recognition corresponds to the size of the image displayed by the projector for the presentation.

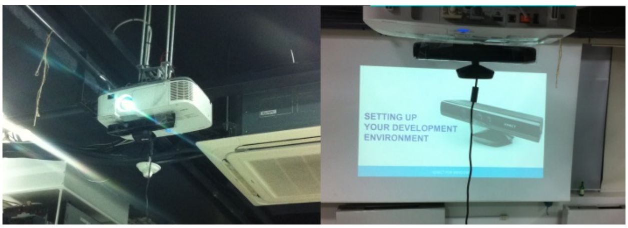


Fig. 1. Experimental space

In the experiment, we can confirm the start and end function of slide show is controlled by voice recognition, the sliding movements that control the next and previous slide, the movements to switch functions, the memo function and eraser function are all controlled using the gesture recognition integrated in our system.

Assuming a situation where the user has a presentation to do, he can starts this presentation by saying "start presentation". While the presentation progresses and the user wants to display the next slide, he puts his hand in a grip posture, moves it from left to right then release his hand posture to finish the gesture. Also, when we want to go back to the previous slide, the same gesture using the left hand and opposite direction is available (Fig. 2).

Fig. 2. Move next Slide & Move previous slide

In addition to the function to change the slides, our system provides functions in order to edit the slide on the fly, such as memo function and erase function. To use them, we must change the cursor directly in the Powerpoint application. To do so, we create gesture to switch from one function to another. When we are in the mode to change the display of the slides, we can pass to the next mode (memo tool) by using a gesture. Similarly we can pass from memo tool to eraser tool, or from eraser tool back to change slide tool. This is the switching function of interface(Fig. 3).

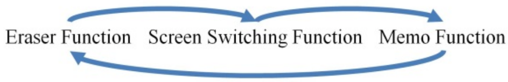


Fig. 3. Switching function root

The memo function is based on the coordinates of the hand. Memo directly modifies the slide. However, if we record all the hand gestures, we cannot obtain exactly what the user wants to draw, because the hand is constantly moving. So we use the grip posture to ease the drawing process. When the user draws with the memo tool, it just has to adopt the grip posture and draw. When the hand is released, the movement of the hands is free and no drawing is made. The same process is used for the eraser tool. When the user is on the eraser function state, it can just grip the hand and move it to erase the content of the slide, and then release its hand to stop erasing.

We chose to use the grip position of the hand to make sure we only consider what the user really want to do, and reduce the errors that are made with undesired gestures. The system is easy to use, since the user just has to change to the desired state of the function he wants to use, then grip his hand to do the gestures(Fig. 4).

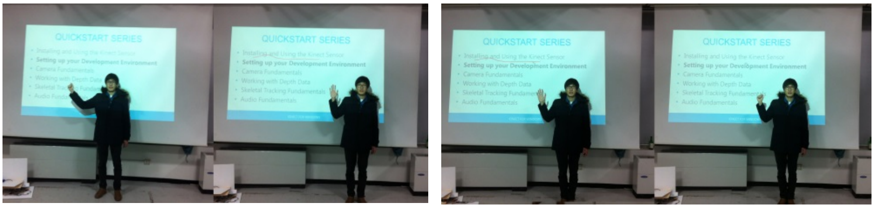


Fig. 4. Memo function & Eraser function

To finish the presentation, the user says "End Presentation". This automatically ends the slide show and the system stops the gestures recognition functions.

5 Conclusion

We introduced an application that makes it possible to start or end the slide show by using the voice recognition through a microphone Kinect. Also, by converting the skeleton image of someone with the Kinect's camera, it is possible to control the presentation. Unlike other researches that just allow to change slides, we added some functions such as a memo tool and an eraser tool. These functions augment the presenter capabilities a lot. Unlike previous studies which were only using the sweep movement, our application allows to use sweep and grip at the same time, and are used to effectively prevent undesired operations from the presenter. It is expected that the voice recognition module and gesture recognition module are used to control a presentation, but they may be re-used for other studies in multiple context, and will be of great help in future studies.

References

1. Bolt, R.A.: Put-that-there: Voice and gesture at the graphics interface, vol. 14(3). ACM (1980)
2. Osunkoya, T., Chern, J.-C.: Gesture-Based Human-Computer-Interaction Using Kinect for Windows Mouse Control and Power Point Presentation
3. Girbacia, F., Butnariu, S.: Development of A Natural User Interface for Intuitive Presentations in Educational Process. In: Conference Proceedings of eLearning and Software for Education, vol. (2) (2012)
4. Ren, Z., et al.: Robust hand gesture recognition with kinect sensor. In: Proceedings of the 19th ACM International Conference on Multimedia. ACM (2011)
5. `http://kinectpowerpoint.codeplex.com`
6. `http://www.samsung.com/sec/consumer/tv-video/tv/`
7. `http://stuffsirisaid.com/`
8. `http://www.samsung.com/global/galaxys3/svoice.html`
9. `http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx`
10. `http://en.wikipedia.org/wiki/Hooking`

Hardware Based Distributive Power Migration and Management Algorithm for Cloud Environment

P. Prakash[1], G. Kousalya[2], Shriram K. Vasudevan[1], and Kawshik K. Rangaraju[1]

[1] Department of Computer Science and Engineering,
Amrita University, Coimbatore, Tamilnadu, India
{npprakash,shriramkv}@gmail.com
[2] Department of Computer Science and Engineering,
Coimbatore Institute of Technology, Coimbatore, Tamilnadu, India
kousir@gmail.com

Abstract. In cloud computing setup of computers power consumption among the distributed computers needs to be minimal with every server running increases the power cost by an average of 50w-100w. A real time implementation of an algorithm to minimize the power consumed in a setup of a parent computer a PIC microprocessor and connected servers is needed to manage the unwanted waste in energy. The usual traditional scheduler doesn't meet the requirements. We program the Microcontroller to implement our algorithm which ensured that minimum number of servers run for a given numbers of virtual machines. The Distributive Power Migration & Management Algorithm for Cloud Environment that uses the resources in an effective and efficient manner ensuring minimal use of power. The proposed algorithm performs computation more efficiently in a scalable cloud computing environment. The results indicate that the algorithm reduces up to 28% of the power consumption to execute services.

Keywords: Cloud Computing, Power saving, Virtual machines, Power migration, Scalability, Microcontroller.

1 Introduction

In the current implementation of cloud technologies with servers and datacenters forming the backbone of its infrastructure is increasing exponentially with the demand from a common user to huge Multinational companies have increased the number of servers and the datacenters all of which need to be powered ,maintained and managed. Constant care is needed to monitor the power consumed by these servers as these play a crucial role not only in its working but also with environmental health as these consume enormous amounts of energy.

The McKinsey report [9] states that the energy consumption of datacenters come around $11.5 billion in 2010 and doubles every five years. Gartner [10] also estimated that worldwide IT infrastructures are responsible for 2% of global CO2 emissions and energy related costs account for the 12% of the total economical expenditures. Also

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_12, © Springer-Verlag Berlin Heidelberg 2014

the huge amount of Carbon-di-Oxide reduces the life time of the servers and datacenters.

Therefore, it's apparent that the need of the hour is methodologies to manage the no of servers that are running which in turn govern the datacenters.

We introduce the Power migration and management algorithm that controls the no of resources that are being utilized. It's a powerful tool in not only running the Virtual machines smoothly but also in mitigating the total power consumption.

2 Related Work

The following are the researches that have been done in the area of interest. One thing is observed that, though lot of research is on in cloud, there are very little initiatives with respects to power management are carried out. And it served as the major motivation factor for us.

C.Gunaratne explained energy level management techniques at the network level. The author provides a method which dynamically changes the link speed of the network [1]. The study also reveals how adaptive link rate (ALR) helps to manage the energy consumption in the Ethernet cards.

ElasticTree [2] a prototype, which animatedly scale up and down the group of live network elements to manage the data center traffic flows. ElasticTree uses both Link State Adoption (LSA) and shutting down the switches which are idle for some time.

In [3], authors use Dynamic Voltage and Frequency Scaling (DVFS) method, to scale up the voltage and frequency of the processor. The dynamic power management techniques are applied into the hardware and firmware levels. Authors also explained set of hardware dynamic power management techniques which is grouped into either Dynamic component deactivation or Dynamic scaling performance.

The Clock network consumes 30-50% of the total dynamic power of the processor. The effective Clock Level Gating [4] can help us to provide dynamic efficient power saving. The power saving was achieved by reducing the voltage and frequency of the components which are idle for some amount of time. The clock level gating is one way of power gating technique used in integrated circuits. Even if the clocks are halted there may be some power leakage. The supply shutdown mechanism is used to overcome the power leakage issue.

Autonomic Virtual Machine Placement in the Data Center [5] gives an abstract level of virtual machine placement in the Data Center. They proposed an autonomic controller which dynamically maps the virtual machine into the physical nodes by satisfying the service level agreement between the provider and the user. The modified Simulated Annealing algorithm is used for their prototype development.
Reputation-based scheduling for dynamic resource mechanism [12] helps to provide a certain level of resource sharing, by considering the available computing nodes in the cloud environment.

In [6], authors reveal two variants of power management in cluster computing such as Static Power Management (SPM), Dynamic Power Management (DPM). Static Power Management methods are applied into hardware components which consumes low power. Based on the acquaintance of the resource consumption and application workloads the Dynamic Power Management techniques are applied.

Lefevre et al. [7] projected a new cloud data center architecture called Green Open Cloud (GOC). The future generation cloud data center Green Open Cloud provides a facility such as prior reservation for the resources. GOC computes the total workload by adjusting the resources with some of the servers which is idle for long time. Green Cloud architecture [8] supports Virtual Machine (VM) migration and VM placement to reduce the energy consumption in the data center.

3 Proposed Solution and Architecture

The below diagrammatic representation (Fig. 1) would be suffice for anyone to understand the underlying idea.

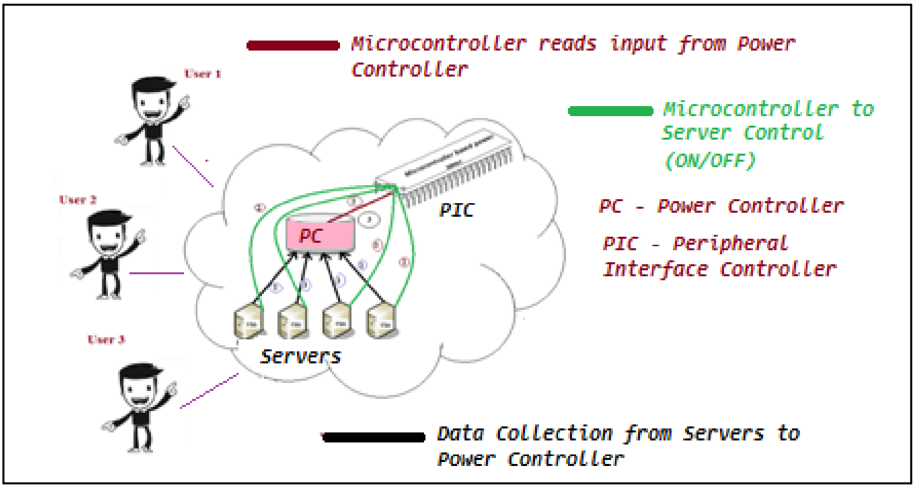


Fig. 1. Proposed Architecture Model

For the benefit of the readers, the following step by step approach is presented along with a clear flow diagram which explains the process accurately. (Refer Fig.2)

Step 1: Power controller get the server details from cloud environment using Distributive power migration and management algorithm.

Step 2: Power controller identifies which server needs to be SHUT DOWN/ START UP based on the details from step1.

Step 3: Send the identified server details from step 3 to PIC micro controller.

Step 4: PIC micro controller shuts down or up the server by passing the respective signal to the specific server.

The next step in the process is to understand the algorithm ported to the microcontroller is presented below in steps which can make life of the reader much easier.

Primary Objective: To manage or allocate the existing virtual machines to servers in such a way that minimal no of severs and in turn least amount of power is used.

- We control the Power usage by controlling the allocation of virtual machines (processes) to servers at each stage.
- Every incoming VM is allocated a server by an arbitrary order defined by the user. Thus a server cannot be left alone unless its maximum capacity is reached.
- Whenever a VM is to be removed from a server the problem here is to check if the VM's in another server can be shifted to this so that the latter can be powered down.
- This is accomplished at every step by retrieving all the VM's and by allocating them again manually therefore each server is filled to its maximum capacity before another server is involved.
- This leaves no chance for any server to be idle or not run at maximum potency.

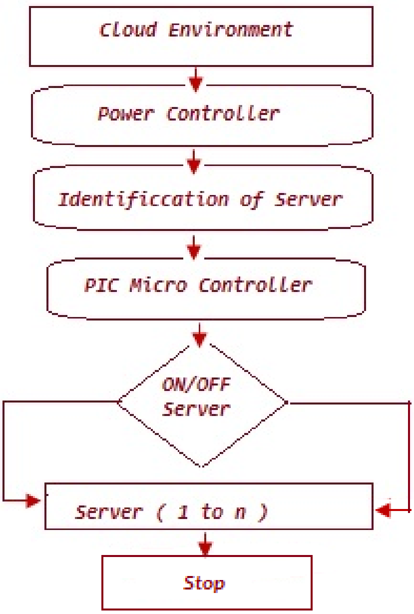


Fig. 2. Flow Diagram

Next step would be to realize the hardware support required for the implementation. This following section will fully concentrate on the same.

The implementation of the algorithm can be seen from the diagram (Refer the Fig. 3). We have a power source with a rating of 12v which is used to power our microcontroller. IC7805 voltage regulator is connected to the power source to ensure consistent power supply to the microcontroller. The microcontroller which contains our algorithm makes use of the available internal clock. If need an external clock, it can be used as in Fig 4 .The algorithm is fed from a USB to the microcontroller to decide which unit to power up or down through the other ports. A maximum of 30 systems can be connected to the microcontroller but due to limited resources we have shown implementation with 6 systems. If needed more systems can be connected using a different microcontroller .Each system unit has a relay which pulls up or down to power on or off the systems.

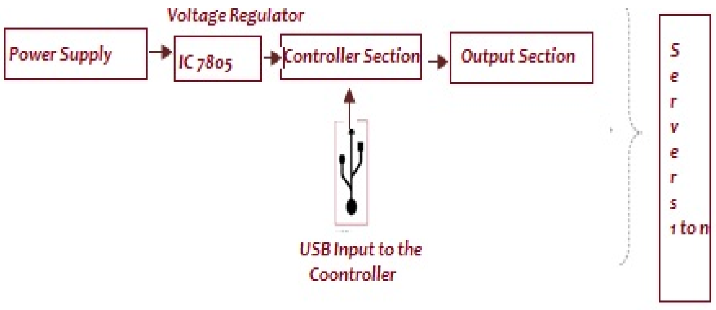


Fig. 3. Overall Hardware Realization

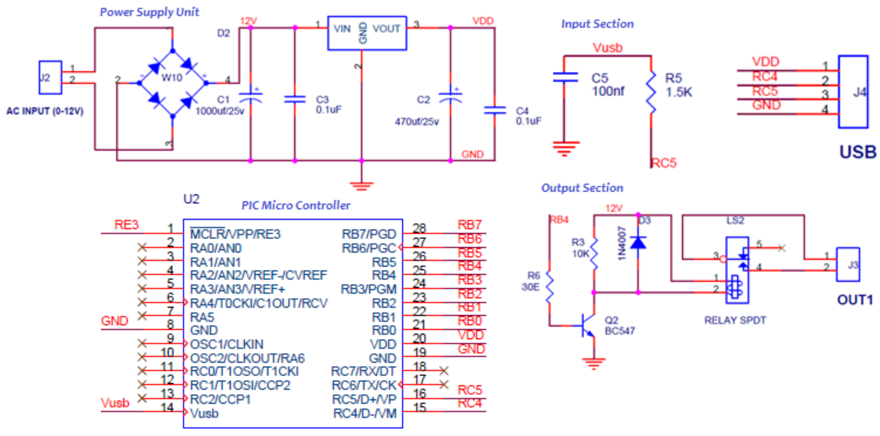
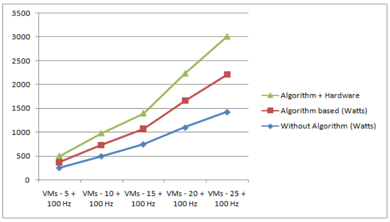


Fig. 4. Expanded Hardware Architecture

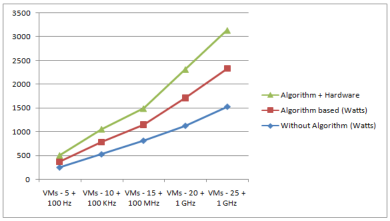
PIC. PIC is Peripheral Interface Controller. PIC architecture is one of the most advanced architecture being used by the industry. 8051 is a basic microcontroller, where it lacked few features as support for inbuilt ADC and DAC, support for I2C or SPI etc., all these have been addressed clearly with PIC microcontroller. PIC is manufactured by Microchip and it is suitable for various and variety of applications. PIC finds its application in mostly automotive and industrial applications. (Kothari et al, 2012).
Here 18F4550 microcontroller is taken as reference. Its high performance can be attributed to many architectural features.

4 Results and Future Prospects

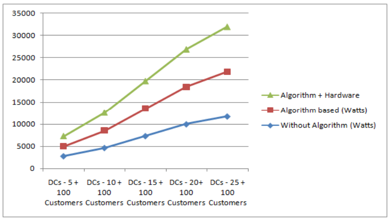
After having a thorough analysis the following graphs are drawn and it shows it all. Studies have been carried out with increase in VMs, increase in VMs with consistent increase in bandwidth, Increase in the number of data centres with constant customer size, increase in the storage size and finally with varied customer needs. All of the above are clearly plotted and presented in the following Fig 5.



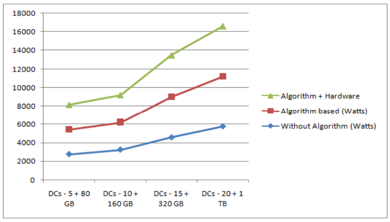
a. With increase in VMs



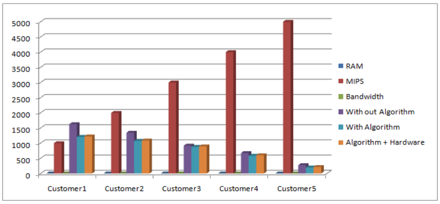
b.With increase in VMs and Bandwidth



C. With increase in Data centres and constant customer numbers



d.Increase in storage capacity



e. With different customer needs

Fig. 5. A Complete Analysis

Most importantly, what can be inferred through this research sounds really interesting. With just implementing the algorithm and when done with simulation, the results are highly satisfactory. When the same has been carried forward with hardware implementation, there found a little lack in the performance. That could be because of latency caused by the usage of hardware circuit. Nevertheless, the algorithm was found to be very effective and it saves lot of power and the core aim in the research has been achieved. Coming to the improvement aspects, the same can be implemented with Zigbee or Bluetooth, wirelessly which would definitely be a milestone in the research in Cloud arena.

Acknowledgements. First author would like to thank second author for providing consistent inputs. Third and fourth authors wish to convey their gratitude to Amrita University for providing an opportunity to use the available resources.

References

1. Gunaratne, C., Christensen, K., Nordman, B., et al.: Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR). Journal of IEEE Trans. Computer 57, 448–461 (2008)
2. Heller, B., Seetharaman, S., Mahadevan, P.: Elastic Tree: Saving Energy in Data Center Networks. In: 7th USENIX Conference on Networked Systems Design and Implementation, Berkeley, USA, pp. 1–17 (2010)
3. Beloglazov, A., Buyya, R., Lee, Y., Zomaya, A.: A Taxonomy and Survey of Energy Efficient Data Centers and Cloud Computing. Journal of Advances in Computers 82, 47–111 (2011)
4. Benini, L., Bogliolo, A., Micheli, G.: A Survey of Design Techniques for System-Level Dynamic Power Management. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 8, 299–316 (2000)
5. Hyser, C., McKee, B., Garner, R., Watson, B.: Autonomic Virtual Machine Placement in the Data Center. HP Laboratories, HPL-2007-189 (2008)
6. Luigi, G., Lassonde, W., Khan, S., Valentini, G., et al.: An Overview of Energy Efficiency Techniques in Cluster Computing Systems. Journal of Cluster Computing (2011)
7. Lefévre, L., Orgerie, A.: Designing and evaluating an energy efficient Cloud. The Journal of Supercomputing 51(3), 352–373 (2010)
8. Liu, L., et al.: GreenCloud: a new architecture for green data center. In: Proc. of 6th International Conference on Autonomic Computing, Barcelona, Spain (2009)
9. Kothari, D.P., Vasudevan, S.K., Subashri, V., Ramachandran, S.: Analysis of Microcontrollers, 1st edn. Scientific International Publishing (2012) ISBN: 9789381714294
10. Kaplan, J., Forrest, W., Kindler, N.: Revolutionizing Data Center Energy Efficiency. Technical report, McKinsey & Company (2008)
11. Gartner Says Energy-Related Costs Account for Approximately 12 Percent of Overall Data Center Expenditures (2011), `http://www.gartner.com/it/page.Jsp?Id=1442113`
12. Hussin, M., Latip, R.: Adaptive resource control mechanism through reputation-based scheduling in heterogeneous distributed systems. J. Comput. Sci. 9, 1661–1668 (2013), doi:10.3844/jcssp.2013.1661.1668

Point Estimation for the Interface Using Acoustic Signal

Jinuk Kim, Jiwoong Heo, Kwangjin Hong, David Jean, and Keechul Jung

School of Media, Soong-sil University, Seoul, South Korea
{jinuk,jiwoong,hongmsz,jeandavid,kcjung}@ssu.ac.kr

Abstract. In this paper, we introduce a new tangible interface that allows someone to hit any parts of a medium such as a wooden block in order to generate original multimedia content. We investigate a new technique to estimate precisely the location in which the interaction occurs using two microphones, Arduino and an amplifier. With this setup, and functions that we describe, we can obtain the time spent from the sound of the hit between the user and the medium, to reach each microphone. These travelling times and the knowledge of the material used as medium are then used to precisely locate the origin of the hit. In the last part of this paper, we show an example of use of our system. We use everyday objects and augment them to provide new unexpected experiences when they are touch by someone. We firmly believe the input technique can be used more generally in many different contexts.

Keywords: Sound Localization, Sound Interface, Acoustic Signal.

1 Introduction

In modern society, the use of computers is ubiquitous and becomes an essential part of our lives. Some decades ago, the interaction with computers was achieved by the input devices provided by their manufacturer, such as keyboard and mouse. However, the price of computer hardware greatly decreased as the use of computers became more popular. Researches in Human Computer Interaction generated the emergence of user-centric computing with new input techniques and devices, such as smartphones using touchscreens. New methods in interaction design focuses on the convenience of the user when using an electronic device, and give to researchers a better knowledge on how humans think and behave.

The advance on human-centered interaction technology in recent years, as well as the rising development of physical computing technologies established a large interest in the field and initiated lots of new experiments. Due to the trend of social interaction, effective input and output devices (I / O Devices) are being developed steadily. The mouse is the most popular input device but is not as intuitive as new methods of interaction that we can see today. Touch screen based devices have emerged to fulfill the lack of visual information and feedbacks of the mouse. Because we are able to directly use the screen without the need to use any other physical devices, we can easily access to the location of the interaction and develop new tools such as touch screen keyboards. There are two techniques of screen touch interaction. One uses

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_13, © Springer-Verlag Berlin Heidelberg 2014

resistive touch screen, and the other uses capacitive touch screen. Resistive touch screen is cheap, accurate and can recognize pressure initiated by nail, pen and even gloves. But, resistive touch screen has some disadvantages. First, the screen has 75% degree of clearness. Second, the reaction time of resistive touch screen is slow. Third, if we don't use it for a long time, it makes the accuracy of resistive touch screen sub-optimal. Lastly, scratches on the touch screen often generate errors.

Capacitive touch screen has strong durability, so the screen performances are not affected by error if they are not used for a long period of time. This make the screen able to recognize touch constantly and it also has short reaction time and high degree of clearness. But the touch screen has to be touched only with hands or skin to recognize the interaction and is difficult to control precisely[1].

There are studies using computer vision about intuitive interaction techniques similar to a touch screen. With hardware development such as infrared camera, some systems using computer vision are able to improve the rate of recognition. But there are disadvantages of using computer vision technologies. For example, light has a strong influence on recognizing object using camera, so lights has a large role on the rate of error. Also computer vision application can only happens in the space covered by the detection range of the camera so the interaction must happen in front of the camera[2,3].

Computer vision is not the only field that made progress in touch based interaction, many studies using sound signal are in progress and give interesting new results. For example, there is a study in which the sound signal is used to control a media player. Making interface that are controlled by sound signal is very recent. Also, there is no hygiene problem if you use this type of interface because you don't need to touch anything. But there is also a disadvantage in this study. The interface processes with only pre-input-signal sounds in a 3-dimension space. The emergence of touch screen has a large influence on IT development. However, the trend to make screen larger and larger has the disadvantage that they can easily be damaged by physical impact[4,5].

In this paper, we investigate a new possible interface that uses acoustic signals as input method. We use a couple of microphones to measure the time to receive specific acoustic signal and compare it with Arduino to estimate the position. The interface we introduce in this paper opens a new approach to create interactive devices and new areas of research.

2 System Structure and Functions

Our system uses 2 microphones. A sound signal is received by the microphones and analog sound signals are converted into digital signal. We are using a specific threshold to create the digital signal while discarding noises. If the sound signal is higher than the threshold, the signal is saved and we can estimate the point of origin. To do so, we compute the distance between each microphone and the origin of the sound. Usually, the signals received by the microphones are too low and it is difficult to recognize them so we use amplifier to improve ratio of recognition. This diagram shows the setup described in this paper (Fig. 1).

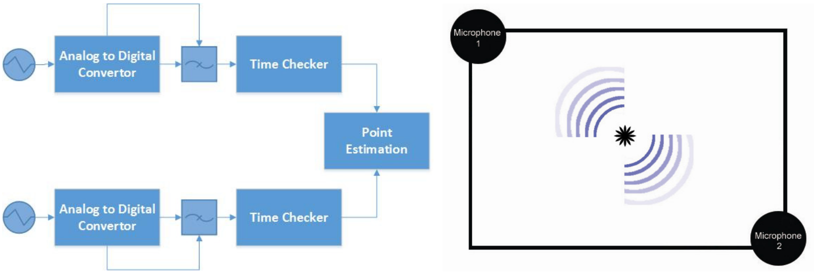


Fig. 1. Interface Structure & Environment of Sound Signal Reception

The system saves and calculates the time of arrival of the sound to microphone in order to estimate the point of origin of the sound signal.

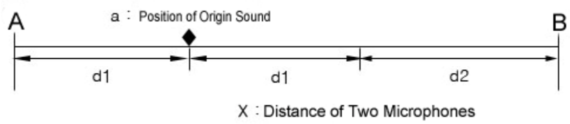


Fig. 2. Method of Distance Estimation

The Fig. 2 above shows the method of estimating a position named ‘a’. If the sound signal is received by the two microphones, each of them gives a time value. IN ou system, we have 2 microphones that we call ‘A’ and ‘B’. We call the time value inputted by the microphone ‘A’ ‘T1’ and other time value inputted by the microphone ‘B’: ‘T2’. We can estimate position of origin sound by using following equation ‘(1)’. :

$$a = \frac{X-K(T2-T1)}{2} \tag{1}$$

K represents the propagation speed of the sound in the medium used ‘(1)’. As you can see in ‘Fig. 2’, ‘A’ is positioned prior to ‘B’. That is because we want ‘T2’ to be smaller than ‘T1’. Then we can subtract ‘T1’ to ‘T2’ and multiply by the sound velocity in order to obtain ‘d2’ then locate the position ‘a’ by subtracting ‘d2’ from the two microphones positions and finally divide the result by 2. The final results we obtain give us the point of origin of the sound[6].

3 Experimental Results

In our experimental setup, we use Arduino Uno. The microphone used is ‘MAX4466 with Adjustable Gain’ equipped with amplifier. The microphones have already an amplifier but the values inputted by the microphones are very small so we use another amplifier.

The distance between the two microphones is 170cm and the amplifiers are fixed on a wooden board. Our main objective is to obtain the time difference between T1 and T2. In order to do so, we need to insure that the 2 microphones are at the same distance with the amplifier. This is to make sure that the travelling time of the signal inside the wire is the same for both signals. Also, in order to detect the sound, this sound needs to be above a certain threshold. A slight and silent touch of the wooden board may not be sufficient to be detected.

We had to process slowly and carefully to do this experiment. We shouldn't experiment constantly, because sound signals in our medium (wooden board) move in wave form. We also minimized the unrelated sounds (noises) except the sound occurred while hitting the board. Many sources of errors can influence the result: the volume of the sound that can be too high or too low, the origin of the sound that must be inside the interactive space of the device, and also the surrounding noise. To minimize the error, we developed a filter within Arduino.

The speed of sound velocity in the air is 340 m/s but the higher is the density of medium, the faster is the speed of the sound. We could not know right away correct sound speed in wooden board used in this experiment so we just started with a simple guess for the first value of 'K' then we could estimate the real value of 'K' with more experiments.

K is a constant that can be drawn through the experiment above (1). During this experiment, we suppose that 'the source of the sound is at the exact same position than one of the microphone'. We start recording the time when a button is pressed and then we save the time that the signal took to reach the second microphone though our wooden medium. We used the time value as denominator and the distance between the microphones and the source as the numerator.

Constant K was estimated using the average time considering the allowable error. The slab of wood used for the experiment was 0.045cm/s and when we converged the propagation speed of the sound inside the slab, we obtain a result of 560m/s, which is faster that the sound speed in air but slower than other mediums. This is due to the wood having less mass that other solid mediums. By applying the constant K to the actual equation (1) we are now able to process to the experiments and compute with accuracy the location of the source of a sound. During the experiment, we sent 50 sound signals from the same distance and calculate the average. The following chart shows the results.

Table 1. Result of Experiment

Real Distance	Difference of Time Value(μs)	Estimation Distance(cm)
40cm	1906.60	42.1013
60cm	1143.13	57.7794
85cm	171.82	81.1339

As we can see in the results, the gap between the real distance and the presumed distance is 2.1013cm for a distance of 40cm, 2.2205cm for a distance of 60cm and 4.8004cm for a distance of 85cm. The next Fig. 3 shows this result in a normal distribution graph.

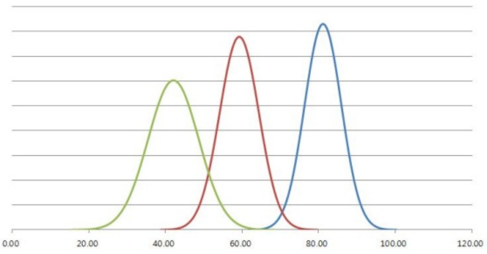


Fig. 3. Normal Distribution Table of Experiment Result

As a result, we figured out that under than 4cm of observable error range, we can recognize the coordinates of the source of sound and we are able to transmit into a database that can be used by any computer programs.

4 Content Creation

In this part, we are going to show that the result we obtained previously allows our system to be a reasonable option to support interactive works. We made a virtual instrument using the interaction method mentioned previously, a wooden board divided into several areas in which the user is able to interact and that generates sounds.

The system consists of Arduino, 2 microphones, a projector and a speaker. The Arduino control the entire system. The wooden board takes form of a bookshelf with the following dimensions: 210cm horizontally and 80cm vertically. We used a bookshelf that can hold books and can be used as a way to interact. Our objective is to give to an everyday object a new dimension as an interactive device.

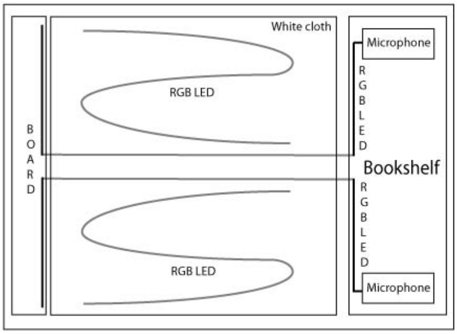


Fig. 4. Content Diagram

If a user taps on any part of the bookshelf, the distance is calculated by the method previously mentioned. A specific sound is then generated corresponding to the tap location.

We divided the bookshelf into 7 distinct areas that are associated with unique sounds or functions. However, nothing indicates to the user the location of each part, as we don't want him to know the boundaries to maintain the idea that this is just a normal bookshelf, although we gave it new capabilities. We want the user to be able to use everyday items, but feel a new experience with them.

Interaction with the bookshelf generates auditory and also visual changes in the environment, each of them unique to specific part in which the user interacts. When the user taps the board, the system progressively modifies the environment, by controlling the playback of a video and also changes the intensity and colors of several light-emitting diodes (LED). At the same time, the auditory effect produced simple digital sounds, also controlled by the location of the hit from the user. Also, the tempo in which the user interact with the bookshelf control both the video playback and the sounds generated. Each bit creates a chord of harmony.

When someone hits first section (0~45cm), some objects appear at first circle line, sounds of drums occurs and the LED corresponding to the hit position turns on, then some specific patterns appears on the screen. If the second section is hit (45~75cm), some objects appear second circle line and user can hear percussions sound and LED is also turned on at the hit position. Different patterns appear on the screen. To hear piano sound, we have to hit third section (75~105cm). A corresponding LED turns on as mentioned above. The Fourth section (105~135cm) pairs with bass band sounds and the fifth section (135~165cm) with electric guitar sounds. Last section (165~210cm) removes all the sounds. All object's movement are linked together with sound's volume and pitch and the resulting patterns are organized to look like a flower. The last part of the bookshelf is used to turn off the system.

By simply and funnily hitting the bookshelf or other everyday items with its hands, the user controls the virtual instrument. Someone passing by may have no idea that a simple wooden board is the support of an entire interactive system. When an interactive system use devices from computer vision techniques, such as Kinect, it is easy to guess what to except when the device is seen. However, this system uses a simple surface to propagate the acoustic signal and it allows us to surprise the user by such an unexpected method. This provides a new experience, making use of a support as an interface that is augmented by new technology and generates fun contents.

Fig. 5. Interface Structure & Environment of Sound Signal Reception

5 Conclusion

In this paper, we introduce a new way to use sound as an input method for an interactive system. We show how to estimates the propagation speed of the signal in a medium, and then use this result to locate the position in which a hit occurs in the medium. Our system uses Arduino to process the acoustic signal by connecting an amplifier to the microphone to receive and process it more accurately. We show all the methods and equation used in order to create our system.

Also we developed an interactive multimedia application to demonstrate the possibility of the input method. We show that using this method allows us to surprise some users and generate new experiences with everyday objects. In the proposed research, there are some issues called by tolerance. That will be complemented by further research. The purpose of this study was to understand and develop new ways to use the propagation of an acoustic signal through a solid medium and use the time to compute précises locations of interaction. We hope that develop the sensible interface through the combination of various materials.

References

1. Lee, J., Jung, W., Shin, Y., Park, S., Jang, D.: Patent analysis for Touch Screen Panel. In: The Proceedings of the 23rd Conference of the Korean Institute of Industrial Engineers, pp. 1275–1280 (2011)
2. Yang, X., Liao, C., Liu, Q.: MixPad: augmenting interactive paper with mice & keyboards for cross-media and fine-grained interaction with documents. In: The Proceedings of the 20th ACM International Conference on Multimedia, pp. 1145–1148 (2012)
3. Sra, M., Lee, A., Pao, S., Jiang, G., Ishii, H.: Point and share: from paper to whiteboard. In: The Proceedings of the 20th ACM International Conference on Multimedia, pp. 23–24 (2012)
4. Stødle, D., Bjørndalen, J.M., Anshus, O.J.: A System for Hybrid Vision- and Sound-Based Interaction with Distal and Proximal Targets on Wall-Sized, High-Resolution Tiled Displays. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 59–68. Springer, Heidelberg (2007)
5. Scott, J., Dragovic, B.: Audio Location: Accurate Low-Cost Location Sensing. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 1–18. Springer, Heidelberg (2005)
6. Jung, S., Kho, N.: Location Express Using a sound's detection. In: Proceeding of the 2010 Conference of the Institute of Control, Robotics and System, pp. 200–202 (2010)
7. Yoon, J., Byun, Y., Choi, K.: The study on time difference measurement of microphones for sound source. Journal of Korean Society of Mechanical Technology 4(1), 1–7 (2002)
8. Kim, J., Kim, D., Yoo, B., Kim, K.: A Study on Sound Source Localization Using a Microphone Array. In: The Proceedings of the 2005 Summer Conference of the Korean Institute of Electrical Engineers, pp. 2753–2755 (2005)

Generic Distributed Sensing in Support of Context Awareness in Ambient Assisted Living

Bin Xiao*, Theo Kanter, and Rahim Rahmani

Department of Computer and Systems Sciences, Stockholm University
Forum 100, Isafjordsgatan 39, SE-164 40 Kista, Sweden
{xbin,kanter,rahim}@dsv.su.se

Abstract. Researches in ambient assisted living have so far faced three important challenges: (1) Lack of a comprehensive approach to capture user needs that are generic; i.e., not limited to specific events, but as generic related to the user. (2) Lack of a highly flexible and scalable platform for the distributed sharing and processing of context between nodes in IoT networks. (3) Increased amount of communication and devices with sensors participating in the acquisition, processing and sharing of context further challenges both computation capability and storage capacity of the system. In this paper, we address these limitations and present novel support, applied in a system for remote assistance of elderly. The support comprehensively retrieves user needs from generic context, via a scalable overlay providing increment of processing capability and storage. Further, the support self-organizes entities into generic context from distributed sensing, using the Dependent Context Pattern (DCP) based on the Context Virtualizing Platform (CVP).

Keywords: Generic Context, Context Dependency, User Needs, Virtualization Platform.

1 Introduction

With the development of global society aging, more and more elderly suffering from slight dementia needs to be cared or supported in daily life. Researchers utilize sensor technology and multi-media to make they live independently, called Ambient Assisted Living (AAL) [1]. In recent research, the AAL service shouldn't just aim at long-term monitoring. Instead, researchers need to focus more on capturing what the users interested in, and provide a dynamic system adaption [2].

Utilizing heterogeneous sensors to detect the context related to users and making an intelligent responding to the users (such as reasoning) are the primary paradigm for contemporary research. However, even sensors have advantages like: objective, easy-to-control, and stable, but conventional physical sensors are not powerful enough to fully capture the flexible and comprehensive context of users [3]. In the conventional view of ambient intelligence, researchers emphasize too much on the physical side of

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_14, © Springer-Verlag Berlin Heidelberg 2014

environment for user context, such as light, temperature, heat, etc.. But the user context in AAL should not be just limited by this, because users live under a quick changing human surroundings with all kinds of events happening. For example, early this morning, the subway is stopped since a terribly heavy snow. But the elderly planned to go outside for shopping. With a slight dementia, she may not know how to deal with that sudden event. Of course, we can expect caregivers to give her a call as reminder, but it is not security-guaranteed. Because the situation is changing quickly with quite many potential problems and some of them not even be recognized by the caregivers. For such case, the "context" we need to acquire is not just the physical status inside the room. Rather than that, we need a broad channel to capture context dispersed in the human surroundings.

On the other hand, with the quickly increasing nodes amount in the system, system computation pressure increased quickly. For this issue, most researchers make compromise between a centralized infrastructure and self-support infrastructure [4]. The centralized infrastructure means system owns a central point to schedule nodes and processes the context aggregately. The self-supported infrastructure indicates that some context should be locally processed by the nodes but the processed results are shared with other peers. In that case, local nodes must be capable enough in some extent for context processing, which introduces a problem of "heavy client". However, a fully distributed and virtualized infrastructure break the impasse of centralized or self-supported, which allows the nodes processing the context in a fully distributed platform, without any central point, without local heavy context processing [5].

The paper will be formed like this. Chapter 2 propose three challenges for AAL research. In Chapter 3, section 3.1 will discuss the Generic Context (GC) sensing and the Dependent Context Pattern (DCP) to solve the challenge one, while section 3.2 will propose the Context Virtual Platform (CVP) to solve challenge two and three. Chapter 4 will carry out model checking. Chapter 5 will draw a conclusion.

2 Research Challenges for AAL

It is an important research problem in AAL to capture user needs and make system adapt to that, in the view of most researchers [6]. For user needs mining, we need a comprehensive and full approach to capture user context. However, most available indoor AAL systems only emphasize on physical factors in the house (e.g. user gesture, light, temperature, etc.) [7] [8] [9]. Partial attention on using physical sensors to acquire context leads some blind area of human surroundings and makes the context limited and inflexible. Except the conventional context for physical environment, some implicit context hiding in human surrounding must be discovered, such as some accidents or disasters happening around the user, weather changing, traffic situation, and even the supermarket discount. With contribution from those contexts the system can capture user needs more humanly and precisely. For instance, with knowing of the coming thunder storm, system can remind the elderly with dementia to wear more clothes or take umbrella before leaving home [10]. By aggregating context from physical environment and human surrounding, we form a Generic Context (GC), which makes the system deduce user needs more precisely and specifically, even acquiring the user social needs. It will be discussed in details later.

However, no matter the physical environment context or the Generic Context (GC, aggregation of physical environment and human surrounding) is highly dispersed and existing with different nature, which is difficult to acquire by only using conventional physical sensors. GC describes the quick changing surroundings using the Dependent Context Pattern (DCP), to which conventional sensors are not flexible enough to track and detect. Moreover, since the context existing discretely, we have to associate each context fragment for context awareness, where the DCP works. Actually, what we should aware of are patterns of the context changing and their relations, which reflecting the real world. As compensation to the physical environment context, human are involved in the system to contribute for human surrounding context. Human is more flexible and capable than physical sensors to precisely capture some context by crowdsourcing. Fig.1 (left) shows the paradigm of retrieving user needs from the GC. Human participation keeps acquiring and sharing the human surrounding context. The context contributed from human is irreplaceable which compensate to the physical environment. By associating things and events retrieved from the GC, user needs can be comprehensively exposed. As a short conclusion, we introduce the first challenge needs to be solved in our research.

Challenge one: Need a comprehensive approach to fully acquire the dispersed and flexible context.

The context processing pressure increase quickly, because more and more context needs to be acquired, and more and more nodes are joining into the system. On the other hand, the context log grows quickly with time, which is very important to be kept for ambient intelligence computation. There are many platforms available for context collection, processing, and storage. But none of them is highly scalable and extensible [11]. Some of them are half-distributed, which introduce a central point for context processing and nodes scheduling. Those non-distributed and semi-distributed architecture makes system difficult to be extend and scaled. There exist two disadvantages for those systems: *(1) Each fault in the central point could be deadly for the system. In quite many cases, a small error in the central point leads a system crash. (2)With the increasing of context waiting for processing, the system is facing with an increasing security risk and computation pressure.* To avoid those disadvantages, we need to keep away from the centralized structure. As a short conclusion, challenge two is introduced as:

Challenge two: need a scalable and extensible platform for retrieving, sharing and processing disperse context.

Some available systems introduce specialized hardware or specialized machine in the nodes for local context acquiring, processing and storing, called “self-supported approach” [4]. They keep part of computation task locally and store some context locally, so as to alleviate the two problems (mentioned above) for the centralized topology. However, that approach puts requirement on context processing capability for the introduced hardware and devices, which may increase the system. As a short conclusion, challenge three is introduced as:

Challenge three: increasing nodes and context log cause heavy context storage and processing pressures on both central point and other distributed nodes.

The Context Virtual Platform (CVP) will be introduced to deal with the challenge two and challenge three, in next chapter. CVP is a fully distributed virtual platform based on the Mediasense, without any central point [11]. With more and more nodes joining into, CVP increase the system computation capability by integrating the resource of each node together, as a virtual platform. Thus the challenge two and three are solved. As the peer-to-peer bidirectional communication basis of CVP, Mediasense is a fully distributed platform for internet of things, which uses Distributed Context eXchange Protocol (DCXP) to share the data between each node seamlessly [11]. Different from other systems, it is lightweight, scalable, and extendable, without central point.

3 Paradigm, Pattern and Platform for Generic Context

3.1 Generic Sensing Approach and Dependent Context Pattern

Human surrounding context includes user profile, surrounding dynamics, and tasks [12]. User profile relates to social relations, user disease history, etc.. Surrounding dynamics means events happening and status related to the user (e.g. weather forecast, disaster forecast, etc.) [12]. Task means activities and goals users should engage into [12]. For the assisted living system, it is important to hold the context of human surrounding. User profile and user tasks play vital roles to decide how "smart" the system can react to user needs, such as health status, habit, and even social relations. With the context from both physical environment and human surroundings, we form the Generic Context (GC) to percept the user needs, as described in Fig.1 (right) [12].

In order to capture human surrounding context, human will participate into the sensing, as shown in Fig.1 (left). Human is more flexible to percept the uncertain and changing context. Also human can diagnose context value and filter some unwanted context [13]. This hybrid sensing approach forms a "human-in-the-loop sensing" [13], by which the system can process context contributed from the human participants by crowdsourcing, then give an attentive and repeatable reminder to the elderly user based on GC. In such way, the first challenge is solved.

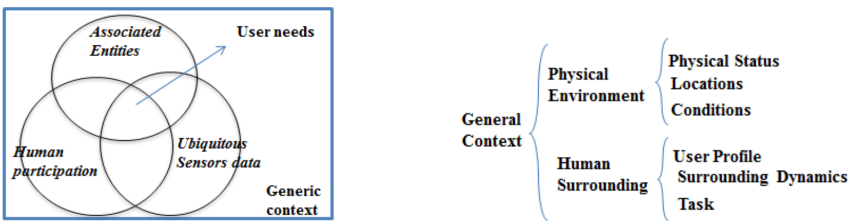


Fig. 1. Generic Context Sensing (left) and Subdivision of Generic Context (right)

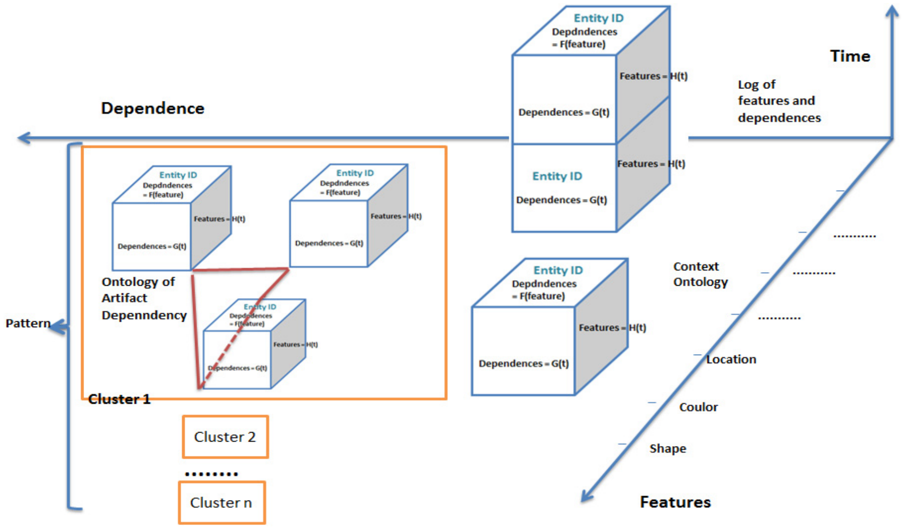


Fig. 2. Ontology based model for Generic Context--Dependent Context Pattern

In this article, we introduce the Dependent Context Pattern to model user context, with three dimensions (i.e. time, dependence, features), by which we define the context entity with a unique ID, composed by group of values on each dimension. Each entity is under certain kind of relations with other entities. Entities with similar features form a cluster, based on which a relation net is established, allowing indexing, query and processing. Fig.2 shows the Dependent Context Pattern (DCP).

The DCP will be used to serve for Generic Context sensing by its five-layer model, where each layer provides service for the upper layer, shown in Fig.3 (right). The observation layer observes physical features of the environment by sensors, which act as system boundary interacting with the world by sensors, such as using Kinect to observe the shape and color of artifacts. Based upon the observation layer, metadata will be delivered to the perception layer, in which features are clustered to form a space of context entity. In the entity layer, system will index and make a matching between the existing entities with their features and the captured features. And create a new entity for the features, who cannot match any existing entity. The semantic layer builds relations among entities by machine learning. With those relations, the context storage is formed. The context layer will output context service using the semantic layer as its input. With the support from DCP, we can comprehensively and fully capture the context for user needs by following a GC sensing way.

3.2 The Virtual Platform for General Context

In this section, we introduce a Context Virtual Platform (CVP) to give an infrastructure support for the layer model DCP, in the view of infrastructure. We try to avoid introducing any central point in the platform to deal with the challenge two and three, by using a completely distributed structure, based on the DCP. In CVP, each node is contributed to process the context on the platform logically, rather than keeping major computation task locally. However, we still hope to carry out the context computation centrally in logic, which benefits for the context flowing and sharing between nodes. Thus CVP provides a virtual platform, which processes the context centrally but distributed in structure. The CVP is a virtualized infrastructure based on Mediasense--an IoT platform which allows bi-directional peer-to-peer communication between nodes [11]. CVP follows the idea of "shared-memory and symmetric multi-processing model" using hypervisors [14]. Based on the Mediasense which provide seamless communication among nodes [11], CVP provides a hypervisor mechanism where each node is capable to join the centrally context processing and cooperated to form a virtual central processor as shown in Fig. 3 (left). In each node, a Raspberry Pi owns its private memory and processor. Under the coordinating of CVP, each node will share part of their private memory and processor with the virtual platform to form a global context storage and a logic-existing central processor. In such way, with more nodes joining into the platform, the virtual storage and virtual processing capability will steady grow, regardless it is in logarithmic increasing or in linear increasing. So CVP makes the system easy to be scaled and extended, with increasing system capability to deal with the increasing system pressure from each node and the central point. Thus, we can solve the challenge two and three by this CVP in support of DCP.

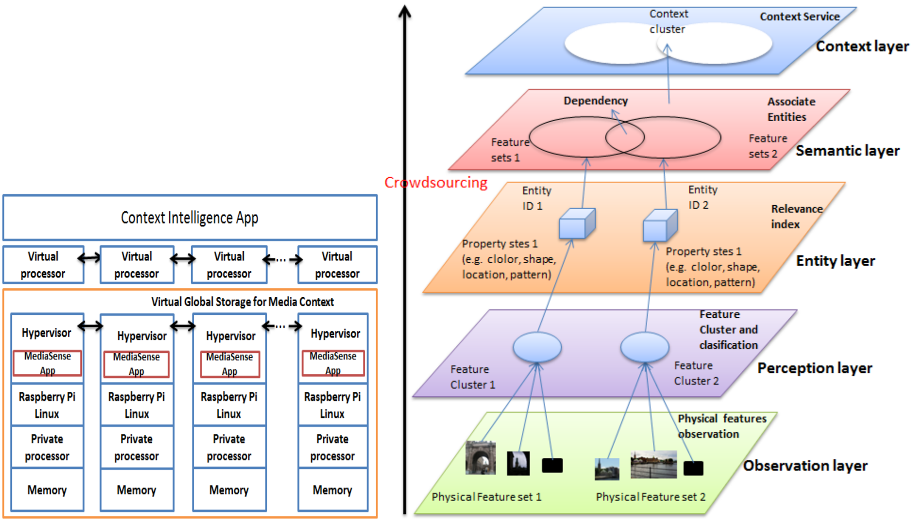


Fig. 3. Context Virtual Platform (left) Supports Layered Dependent Context Pattern (right)

4 Evaluation

We use model checking to verify logic correctness of the proposed models, through the transition systems description for each layer in the layered DCP model [15]. The DCP layer model hybrids process logic in CVP to support the DCP. The transition system is defined as $M=(S,\rightarrow,L)$ [16]. S is a group of states, $\rightarrow$ is the relations between states, and L is a labeling function [16]. Then we declare the system properties φ for each layer and check whether $M \models \varphi$. The labeling algorithm is used in this model checking, where the defined transition models and properties are input then states satisfying with φ are output [17]. Computational Tree Logic (CTL) is used to describe properties of the branching system. CTL provides a semantic to describe both the time features and system path. "A" and "E" are path variables, while "X", "F", "G", and "U" are time variables [16]. "A" means for all the paths, "E" means existing a path. "X" means in the next state, "F" means in the future, "G" means globally in the future, "U" means until. System properties are expressed, using the combination of path variables and time variable.

Table 1. Properties and Proposition of model checking

Layer	Properties and Proposition Formulas
Observe	*{D:distiguished, F:feature-seperated, A:active, P: positive}* *φ 1: AG((D→EF F) & EU(P,A)); φ2: AG(!(P → EF A)& !(F → EF D)).*
Percept	*{Cla:calssification, Clu:cluster-processed, Spe:special-processing, A: active, P: positive} φ 3: AG(P→EX A);* *φ 4: AG (!(A →EF P) & !(Spe & A →EF (Clu∣Cla))).*
Object	*{Exi: existed object(otherwise need a new object); TF: tagged feature, A: active, P: positive;}φ 5: AG(Exi & P→EF (TF&A));* *φ 6: AG((TF&A) → Ø).*
Semantic	*Ass: associated, In: index marked, A: active, P: postive;* *φ 7 : AG((Ass ∣ P→EF In & A) &(P →EF A));*
Context	*C: context, Ser:serving, K: knowledge, A: active, P: postive* *φ 8: AG (!Ø→ EF (S & A)); φ 9: AG !(S →EF (K ∣ C)).*

The properties and propositions for each layer are defined in Table 1. The properties (φ) are defined based on system robustness, liveness, and accuracy. Propositions (like *Exi, D, F etc.*) indicate the amount of S and features for each S in each layer. We use the NuSMV as our model checking tool. NuSMV is a symbolic tool, based on SMV [18]. We describe each layer in the layered DCP using SMV model language [18], according to the propositions defined in Table 1. Then we input all the properties in Table 1, to pick out S_c where $M \models \varphi$, until $S_c \Leftrightarrow S$. So we ensure the model quality and logic correctness by controlling the robustness, liveness, and accuracy.

5 Conclusion

The contribution of this paper starts at chapter 2, by defining three research challenges in AAL, which has no enough attention in current research. Aiming at the first challenge, we proposed DCP with GC sensing in Section 3.1. By integrating DCP with GC sensing, we can fully capture the user needs in a comprehensive way, where many researchers did not pay enough attention to in previous research. In Section 3.2, we aim at challenge two and three to introduce the CVP, which supports DCP in infrastructure and form a highly scalable and flexible system for AAL by peer-to-peer communication. It deals with the challenge two and three, where very few researchers have formally provide specific or practical solution. By the prototype, we find our models can well fit for the AAL scenario case. Chapter 4 verifies the logic correctness for DCP and CVP, using Transition System and Computation Tree Logic. By the GC and DCP, we successfully expand the context acquiring channel for previous research. CVP deals with the embarrassing situation of compromise between centralized structure and self-support structure, by providing a flexible and scalable virtual platform with increasing processing capability and storage capacity, where few really practical solution are provided in previous research. Our current work focuses on defining the paradigm of DCP, CVP model, and Generic Context Sensing methodology. The future work is to create an implemented CVP prototype and DCP algorithm.

Acknowledgement. Research funding from the European research project–Smart Assisted Living Involving Informal Care Givers (SALIG++) is deeply appreciated.

References

1. Chiriac, S., Saurer, B.R.: An Ambient Assisted Living Monitoring System for Activity Recognition - Results from the First Evaluation Stages. In: Fifth German AAL Congress, Berlin, Germany (2012)
2. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: technologies, applications, and opportunities. Pervasive and Mobile Computing 5, 277–298 (2009)
3. Boulos, M.N.K., et al.: Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. International Journal of Health Geographics 10(1), 67 (2011)
4. Guo, B., et al.: From the internet of things to embedded intelligence. World Wide Web Journal, 1–22 (September 2012)
5. Fitton, D., Sundramoorthy, V., Kortuem, G., Brown, J., Efstratiou, C., Finney, J.: Exploring the Design of Pay-Per-Use Objects in the Construction Domain. In: Roggen, D., Lombriser, C., Tröster, G., Kortuem, G., Havinga, P. (eds.) EuroSSC 2008. LNCS, vol. 5279, pp. 192–205. Springer, Heidelberg (2008)
6. Emiliani, P.L., Stephanidis, C.: Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM Systems Journal 44(3), 605–619 (2005)
7. Sit, G.F., et al.: Application-Oriented Fusion and Aggregation of Sensor Data. In: Ambient Assisted Living, pp. 3–13. Springer, Heidelberg (2012)
8. Storf, H., Kleinberger, T., Becker, M., Schmitt, M., Bomarius, F., Prueckner, S.: An event-driven approach to activity recognition in ambient assisted living. In: Tscheligi, M., de Ruyter, B., Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) AmI 2009. LNCS, vol. 5859, pp. 123–132. Springer, Heidelberg (2009)
9. Becker, M., et al.: Approaching ambient intelligent home care systems. In: Pervasive Health Conference and Workshops. IEEE (2006)

10. Metso, A., et al.: Ubiquitous Services for Senior Citizens-Living Process Detection and Middleware. In: Eighth IEEE/ACIS International Conference on Computer and Information Science, ICIS 2009. IEEE (2009)
11. Kanter, T., et al.: Mediasense–an internet of things platform for scalable and decentralized context sharing and control (2012)
12. Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to context than location. Computers & Graphics 23(6), 893–901 (1999)
13. Sheth, A.: Citizen Sensing, Social Signals, and Enriching Human Experience. IEEE Internet Computing 13(14), 80–85 (2009)
14. Sobania, J.-A.: Memory Management in a Many-Core Distributed Hypervisor. Ph. D. Retreat of the HPI Research School an Service-oriented Systems Engineering, 165
15. Wang, J., He, X., Deng, Y.: Introducing software architecture specification and analysis in SAM through an example. Information and Software Technology 41(7), 451–467 (1999)
16. Reeves, S., Clarke, M.: Logic for computer science. Addison-Wesley, Wokingham (1990)
17. Baier, C., Katoen, J.-P.: Principles of model checking, vol. 26202649. MIT Press, Cambridge (2008)
18. Cimatti, A., Clarke, E., Giunchiglia, E., Giunchiglia, F., Pistore, M., Roveri, M., Sebastiani, R., Tacchella, A.: NuSMV 2: An openSource tool for symbolic model checking. In: Brinksma, E., Larsen, K.G., et al. (eds.) CAV 2002. LNCS, vol. 2404, pp. 359–364. Springer, Heidelberg (2002)

RESTful Information Resource Pool Based on Cloud Computing

Shaoqiu Shi, Hongming Cai, and Lihong Jiang

School of Software, Shanghai Jiao Tong University, Shanghai, China, 200240
shishaoqiu@126.com,
{cai-hm,jiang-lh}@cs.sjtu.edu.cn

Abstract. With the rapid growth of the complexity and flexibility of enterprise business process as well as the popularity of cloud computing, how to access the resource involved in enterprise business process flexibly is a thorny issue. The existing approaches of enterprise business process have a tight relationship with the data and resource involved, which may cause either overload or waste of resources. In this paper, Resource Pool is proposed, which provides a flexible resource access service for business process in order to maximize the usage of resource in the multi-tenant environment. In order to describe Resource Pool in a formalized way, modelling research is put forward. Through virtualization of entity resources and service, and monitoring of the resource pool instance, the life cycle of entity resources and resource pool instances can be maintained so as to adjust the allocation of computing resources and storage resources in time and to maximize the reusability. Automatically scaling out is implemented to realize flexible access service.

Keywords: Cloud Computing, Business Process, RESTful Architecture, Resource Pool.

1 Introduction

For enterprise information system, business process is no doubt the core. With the trend of cloud computing, how to combine enterprise business processes with cloud platform is a big problem.There are multiple tenants on one cloud platform. Each tenant needs to access different business processes. Each process interacts with its data differently. However, the existing approaches of enterprise business process have a tight relationship with the data and resource involved, which may cause either overload or waste of resources. How to provide a proper data access service for business process in the multi-tenant environment becomes the key problem of our approach.

In this paper, we design and realize an approach of resource pool that is based on RESTful service and business process modelling. The resource pool manages to decouple business process from service and service from data so as to virtualize resource and service involved in business process. Moreover, based on the virtualization, it provides a scalable resource access service in order to maximize the usage of resource in the multi-tenant environment.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_15, © Springer-Verlag Berlin Heidelberg 2014

2 Related Work

In the paper[1], the author proposed Resource Oriented Architecture(ROA), whose main idea is to use REST[2] Web Service to handle business requests. Other than SOAP Web Service which is operation-oriented. REST Web Service is resource-oriented and it's suitable to enterprise information systems, which is also resource-oriented.

The article[3] puts forward the concept of the RESTful Resource Pool, which is resource-oriented and flexible business process solutions. It can adapt the frequently-changed business processes. However, it can’t provide scalable data access services for following reasons: Firstly, it is a bottom-up solution. The databases and tables are built in advance and RESTful services are provided to access these databases and tables; Secondly, there is only one instance of the resource pool and it can easily become the bottleneck of the system.

A BPMN modelling method based on REST was presented in paper[4] that combines BPMN and REST. Designer should add tags to the activities in the process so as to map them with REST web services.

Article [5] presents a flexible resource model for High Performance Computing as a Service in cloud computing, along with a new service scheduling model for multi-tenant cloud resource sharing. While author of [6] designed and implemented a workflow system on academic cloud and introduced a measurement of performance.

These articles and works explore resource data access from the perspective of concept, modelling and implementation. They can solve part of the problem, but can’t completely meet the our requirements. First, they haven't come up with a complete model system; Second, their research isn't from business processes and the entity resources involved, can’t meet the demand of the data access for business processes; Finally, they haven't realize flexible access to resources in the multi-tenant environment. In this paper, we proposed Resource Pool to solve the problems above. It's a scalable business process resource access service.

3 Overall Framework

Resource Pool is the middleware that is responsible for connecting enterprise information systems and applications with data storage of resources. With unified REST service, it provides a flexible and scalable resource access mechanism to upper applications in multi-tenant environment.

In order to provide flexible and scalable services, we should decouple business process from service and service from data so as to virtualize resource and service involved in business process. The virtualization includes entity resource virtualization and service virtualization.

After the virtualization, we can monitor the status the servers to maintain the life cycle of the entity resource as well as the resource pool instance so as to adjust the allocation of computing resources and storage resources in time and to maximize the reusability. Moreover, the resource pool instances are monitored so that they can be automatically scaling out or in according to defined thresholds to realize scalable data access.

4 Resource Pool Modelling

Models of resource pool include static models and dynamic ones (Figure 1). Dynamic models are responsible for linking static ones.

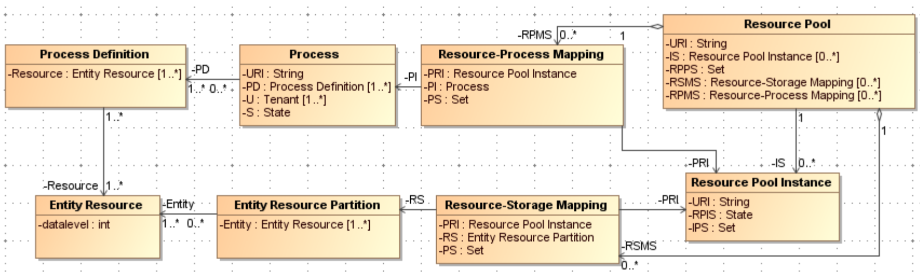


Fig. 1. Class diagram of the models

4.1 Static Models

- **Process Definition** is the formal definition of enterprise business process as models. According to the concept of REST, each resource is represented by a unique Uniform Resource Identifier (URI). A set of entity resources and state transition rules are also included in the processes definition.
- **Process** is the executing instance of certain process definition. Each process has an URI. A process belongs to a particular tenant.
- **Resource Pool** is the container of resource pool instances that is responsible for analysing processes and maintaining resources-process mapping and resources-storage mapping.
- **Resource Pool Instance** is mapped to corresponding process. The states of a resource pool instance include initial, deployed, running and suspended.
- **Entity Resource Partition** contains a set of entity resources that are stored together. Some entities are often accessed together (Flight and Ticket for instance). These entities can be formed as an entity resource partition.

4.2 Dynamic Models

- **Resource-Process Mapping** is the mapping of resource pool instance and process. They are maintained by the resource pool. The process access resources through the resource pool instance.
- **Resources-Storage Mapping** is the mapping of the resource pool instance and the set of entity resources, maintained by the resource pool. The mapping includes a set of properties that indicate the access rule to the storage.

5 Proposed Research

5.1 Virtualization

1. Entity Resource Virtualization After a tenant starts a business process: a) The resource pool parses the entity resources after analysing the process definition. It then create databases and tables for the entity resources according to their data levels(which will be described in the following section). b) Two configuration files are then generated which are a mapping between the process and the resource pool instance and a mapping between the resource pool instance and the storage. c) The REST service then is generated, packaged with the configuration files and deployed as a resource pool instance. By checking two mappings, the resource pool can locate certain resource pool instance as well as the storage it is mapping to.
2. Service Virtualization Each resource pool instance is a set of REST services and configuration files that are packaged together and deployed on distributed servers. We distribute requests in two levels: a) One resource pool instance is created for each process. b) Each resource pool instance is supported by a cluster with certain load balance strategy.

5.2 Life Cycle Management

1. Managing Resource Pool Instance The states of resource pool instance include initialled, deployed, running and suspended. When a process starts, the resource pool instance is created and then deployed. In the running state, the resource pool instance manages to parse the request from application and returns the data back to the application after executing queries on the storage. When a process is terminated, the instance will be suspended and put into the buffer pool.
2. Managing Entity Resource Entity resource has different data levels that lead to different life cycles, including process level, tenant level, and global. a) Entity resource in process level is totally isolated. The database and table will be created dynamically when the process is started. When the process is terminated, the data will be migrated to data warehouse before databases are dropped. b)Entity resources in tenant level are partially shared within the same tenant. The databases and tables will not be deleted until the tenant is removed. c) Global data is permanent and they are shared among tenants. They are usually some public entities, such as the map information.

5.3 Scaling Out

By monitoring the resource pool instance and each node in it, the running status can be retrieved, including memory usage, concurrent visits as well as average response time. The thresholds can be set on these factors with certain response actions including alerting and the scaling out strategy. For instance, when the concurrent visits is over 100, the cluster will scale out automatically.

6 Implementation

6.1 Architecture of Resource Pool

The architecture of the resource pool is shown in figure 2. The resource pool consists of naming node, resource pool instance container, storage module and buffer pool.

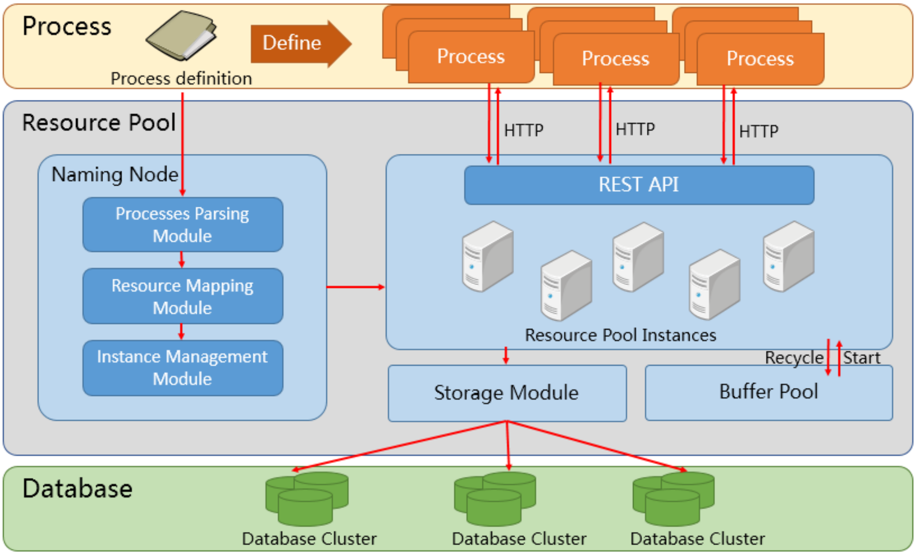


Fig. 2. Architecture of the resource pool

- **Naming node** consists of: a) Process parsing module extracts entity resources and state transition sequence from process definition; b) Resource mapping module establishes and maintains resources - process mapping and resource - storage mapping; c) Instance management module accepts applications' query and returns the location of certain resource pool instance dynamically.
- **Resource pool instance container** maintains the deployed resource pool instances. The buffer pool is responsible for buffering several suspended resource pool instances and redeploying them if needed.
- **Storage module** includes SQLGenerator and the database connection pool. SQLGenerator automatically generates the SQL statement to create databases and tables; Database connection pool is responsible for the allocation and management of the database connections.

6.2 Comparison

In this section, we will compare the traditional solution, other resource pool solution and our solution. Traditionally, the resource access services are modelled as activity nodes in process definition. These services are created, deployed and started before the process is started[1]. The other resource pool is the approach presented in [3]. The result is shown in figre 3.

		Traditional solution[1]	Other resource pool[3]	Our resource pool
Implementation	System Implementation	Large amount of coding	Small amount of coding	Small amount of coding
	System maintenance	The dependency on the original programmer is heavy	When the model of enterprise is changed it has to regenerate the whole resource pool	The change of the models will only affect the only one resource pool instance
Management		Not have	proposed a solution of configuration and management based on resource	A completely life cycle management and monitoring and scaling out
Effect	Flexible access	Not have	Not have	provide a flexible business process data access service
	Cloud platform	Inadaptable	Inadaptable	well adapt to the cloud platform

Fig. 3. Comparison

7 Conclusion

This paper proposes a flexible business process resource access solution based on cloud platform. Firstly, we introduce the overall plan of the resource pool, and model the system from two aspects: static and dynamic. We then expound the relationship between the models. After that, we put forward the architecture and implementation of the resource pool. Finally we compare the resource pool with traditional solutions and other similar resource pools in detail. The result shows that the resource pool we proposed can meet the need of providing flexible REST resource access services for enterprise business process on cloud platform.

References

1. Overdick, H., Hasso Plattner Inst.: The Resource-Oriented Architecture. In: 2007 IEEE Congress on Services, July 9-13 (2007)
2. Roy, T.: Fielding, Architectural styles and the design of network-based software architectures. PhD thesis, Dept. of Information and Computer Science, University of California, Irvine (2000)
3. Wang, S.: RESTful Information Resource Pool Based on Flexible Business Process. Microcomputer Applications 2 (2013)
4. Pautasso, C.: BPMN for REST. In: Dijkman, R., Hofstetter, J., Koehler, J. (eds.) BPMN 2011. LNBIP, vol. 95, pp. 74–87. Springer, Heidelberg (2011)
5. Zhao, H., Li, X.: Designing Flexible Resource Rental Models for Implementing HPC-as-a-Service in Cloud. In: IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum (2012)
6. Gupta, A., Kumar, J.: Design and Implementation of the Workflow of an Academic Cloud. Indian Institute of Technology, Delhi (2012)

An Embedded Control System Designed Based on Soft PLC

Qingguo Zhou[1,*], Xuhui Yang [1,2], Genliang Han[2], Jiaming Yang[1], Guohan Liu[2], Lei Yang[3], and Ji Qi[3]

[1] School of Information Science & Engineering, Lanzhou University, Lanzhou, 730000, China
zhouqg@lzu.edu
[2] Institute of Sensor Technology, Gansu Academy of Sciences, Lanzhou, 730000, China
[3] Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, Gansu, China

Abstract. With the development of computer technology and established of IEC6113-3 standard， the development of soft PLC technology is strongly developed. Compared with the traditional PLC, Embedded control system based on soft PLC provides the advantage of powerful ability of data process, net communication and open system structure, which can meet the demands of the structure, price, power dissipation and portability of control system in the industry automation field. Based on this, the paper designed an embedded control system based on soft PLC. In the second part of the paper we divided the overall system architecture into hardware and software for detail introduce. Especially, it expounded the design concept of the extension control module of the hardware circuit and the implementation method of the soft PLC. the system has the advantages of compact structure, flexible configuration, good portability, high universality, it can widely used in small and medium-sized industrial control field.

Keywords: Embedded System, Soft PLC, Control System, IEC6113-3, IO Extension.

1 Introduction

With the limitations of traditional PLC systems in a variety of applications has become increasingly prominent, in recent years, the soft PLC control system based on industrial PC has been greatly developed in the industrial control [1-3], due to its powerful data processing capabilities, powerful network communications capabilities, better stability and open architecture.

But it still can't satisfy the modern needs of small medium-scale industrial control system [4]. Embedded system, a new computer system, has been widely used in the current market because of high reliability, small size, low power, low cost, etc [5]. If we can integrate features of the embedded system and advantages of the soft PLC into industrial control system, compared to traditional control systems, the control system

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_16, © Springer-Verlag Berlin Heidelberg 2014

will be greatly improved in size, structure, power, portability, and it just to meet the needs of small and medium-scale industrial control.

2 The Overall System Architecture of the Embedded Control System

The embedded control system is organized by the central control module and extension control module. The two parts are connected via the system bus. Figure 1 shows the system architecture.

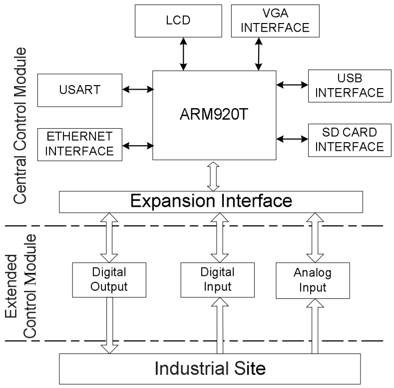


Fig. 1. The system architecture of the soft PLC

The central control module is the core module to execute data processing and control tasks, and it is also the execution platform of the soft PLC.

The extension control module consist of analog input module and digital I/O modules which can be cut and extension according to user equirements, and it can completes the analog input and digital I/O work.

The soft PLC is a software system which works on the central control module, its main function is to explain and execute user's PLC program.

2.1 Designed of the System Hardware

2.1.1 The Central Control Module

The central control module based on ARM920T CPU core, its basic frequency is up to 203MHz, and it supports for mouse and keyboard input, provides a serial port, USB interface, SD card interface, VGA interface, the extension interface, the Ethernet interface and a 10 bit A/D converter channel. The module's design method is discussed in detail in many references [6], so it will be not introduced in this paper. In order to provide the maximize system resources for control function, improve the system's real-time and reduce system power consumption, the module's software and hardware can be cut according to user requirements.

2.1.2 Designed of the Hardware of the Extension Control Module

Figure 2 shows the architecture of the extension control module, which uses the "Extension base board + Extension cards" model.

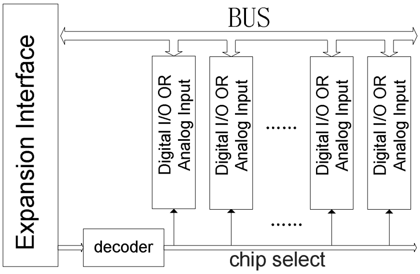


Fig. 2. The structure of the extension control module

Decoder and a number of identical parallel slots are designed on the extension base board, and we designed 8 slots in this paper. The decoder using the 3-8 line decoder 74LS138 chips to complete eight extension cards' selection; the slots are used to connect the extension cards; system bus and A/D conversion channel parallel to eight slots.

The extension cards have digital I/O cards and analog input cards two ways. Both cards connect to the extension base board through a unified interface, it also provides unified input and output interface, and allows user choose the number and type of the extension card according to their requirements. So the design method enhances the flexibility of the system.

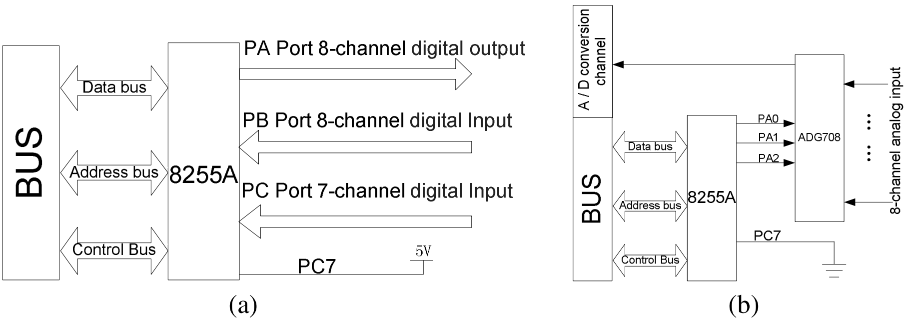


Fig. 3. Schematic circuit of the digital I/O card and the analog input card

The schematic circuit of the digital I/O card is shown in Figure 3(a). Digital I/O card completes digital signal acquisition and industrial field control. We achieve it by using a 8255A chip control relays. PC7 pin of 8255A chip is often high level to mark the digital I/O card (PC7 connect to VCC). PA port is used as 8-channel digital output pins; PB port and PC0-PC6 pins are used as 15-channel digital input pins.

Analog input card realized the collection of industrial field multi-channel analog signal by extending single A/D conversion channel of the central control module into multiple channels. In order to provide a unified interface with digital I/O cards, we uses PA0-PA2 pins of 8255A chip control single 8-to-1 multiplexer ADG708 chip on analog input card. This method can achieve 8-channel analog input. PC7 pin of the 8255A is often low level to mark the analog input card. Figure 3(b) shows the schematic circuit of the analog input card.

2.2 Designed of the System Software

2.2.1 Development Environment of the Soft PLC

We designed the soft PLC based on Windows CE 5.0 operating system which developed by Microsoft. Windows CE is a multi-tasking operating system. It has many advantages, such as highly modular, multiple hardware platforms support, good user-interface, good real-time, etc. We can customize the Windows CE system with Platform Builder which is an integrated development tool. We developed the soft PLC system and the extension card control program using the cross-development environment which called Embedded Visual C++.

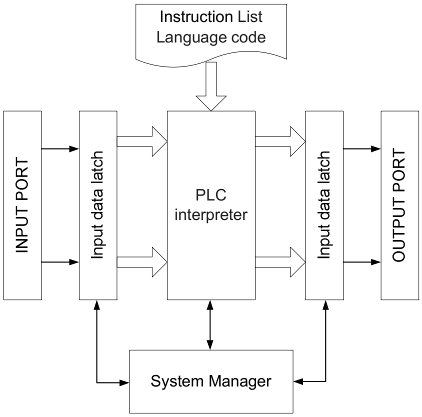


Fig. 4. Functional block diagram of the soft PLC interpreter

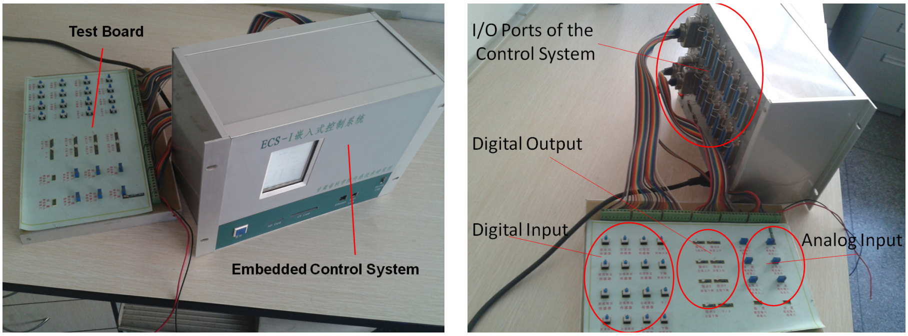
2.2.2 Designed of the Soft PLC

In our previous work, we have implemented a kind of embedded soft PLC [7], the soft PLC using the "Compile + Interpret" scheme. That is, as first to compile user written code into instruction list (IL) code, and then to execute IL code through the interpretation of PLC system.

If we start the PLC interpreter when we get the IL code, automatic control will be achieved by calling and interpreting the IL code. Figure 5 shows the functional block diagram of the soft PLC interpreter.

3 Implementation and Test of the Prototype System

As above mentioned, we designed a prototype system. In the prototype system, we designed 8 slots on extension base board which can insert digital I/O cards or analog input cards randomly. And we designed 8 same interfaces to connect device in the industrial site. It contains one serial port, two USB port, one SD card port and a liquid crystal display (LCD). Figure 6 shows the prototype system and the test board we designed, (a) is a front view of the prototype system and (b) is a side view of the prototype system. In figure 6(a), the test board was used to simulate the industrial field.



(a) (b)

Fig. 5. The physical map of the prototype system

We use a typical PLC program to test the prototype system. Figure 7 shows the ladder diagram of the PLC program. First, we edit the IL program in the PLC editor; Then, we start the PLC interpreter, configure the parameters and load the IL code; At last, the system will automatic control the I/O port according the IL program.

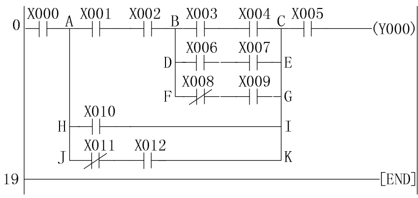


Fig. 6. The Ladder Diagram of a typical PLC program [8]

In this case, the program has 13-channel digital input signal and 1 channel digital output signal. We simulate control components with 13 switches and simulate the controlled component with 1 LED lamp. The results are exactly same with theoretical results and the control system can implement the industrial control.

4 Conclusion and Future Work

The paper designed an embedded control system based on soft PLC which combines soft PLC technology and embedded technology. The system not only inherits the advantages of the soft PLC, but also has the characteristics of the embedded system. The system has strong data processing and control functions with its compact structure, flexible configuration, good portability, high universality, and it can widely used in small and medium-scale industrial control.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant No. 60973137, Program for New Century Excellent Talents in University under Grant No. NCET-12-0250, Gansu Sci.&Tech. Program under Grant No. 1104GKCA049, 1204GKCA061, and No1304GKCA018. The Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2013-k05, lzujbky-2013-43, lzujbky-2013-44 and lzujbky-2012-44, Google Research Awards and Google Faculty Award, China.

References

1. Katz, R., Min, B.K., Pasek, Z.: Open architecture control technology trends. ERC/RMS Report 35, 35–38 (2000)
2. Koziorek, J., Landryova, L.: Softplcs and their using in control applications. In: The 5th IFAC Workshop on Programmable Devices And Systems (2001)
3. Designer's Guide of Proview. SSAB Oxelösund AB (2006)
4. Chmiel, M., Hrynkiewice, E., Milik, A.: A New Compact Programmable Logic Controller with Integrated Programming Equipment. In: The 5th IFAC Workshop on Programmable Devices And Systems, pp. 101–106 (2001)
5. Magro, M.C., Pinceti, P.: Measuring real time performances of PC-based industrial control systems. In: IEEE Conference on Emerging Technologies and Factory Automation, ETFA, pp. 540–547. IEEE (2007)
6. ARM920T Technical Reference Manual. ARM (2000)
7. Yang, X.H., Xi, C.H., Xu, W.D., et al.: Design of embedded soft PLC based on Windows CE. Computer Engineering and Design 24, 026 (2008)
8. Zhu, W., Wang, W., Ding, H., et al.: Open Architecture Soft PLC Based on Embedded PC. Machinery & Electronics 3, 3–7 (2002)
9. Chen, S.L., Chen, I., Shao, H.C., et al.: Development of a Control System with Soft Programmable Logic Controller (I/O) Compiling Function Embedded and Applying into Automatic Optical Inspection System. Advanced Science Letters 8(1), 1–6 (2012)
10. Keeley, T.M.: Industrial control system providing remote execution of graphical utility programs: U.S. Patent 6,138,174 (October 24, 2000)

A Network Coding Based Multicast Routing in Mobile Ad-hoc Wireless Networks

Le The Dung[1] and Beongku An[2]

[1] Dept. of Electronics & Computer Engineering in Graduate School, Hongik University, Korea
thedung_hcmut@yahoo.com
[2] Dept. of Computer & Information Communications Engineering, Hongik University, Korea
beongku@hongik.ac.kr

Abstract. Nowadays, network coding is an emerging approach used to increase network throughput. However, in mobile ad-hoc wireless networks, due to dynamic network topology, it is difficult to make coding decision at intermediate nodes. In this paper, we discuss how to increase coding opportunities and the challenges of network coding in mobile ad-hoc wireless networks. From those issues, we propose a Network Coding based Multicast Routing, called NCMR, in mobile ad-hoc wireless networks. NCMR is especially well-suited for providing reliable, high speed multimedia applications. In our study, the performance of our proposed NCMR is evaluated in terms of packet delivery ratio, end-to-end delay, network congestion via different scenarios by using OPNET.

Keywords: mobile ad-hoc wireless networks, multicast routing, linear network coding, decoding failure.

1 Introduction

In the field of information theory, there are many information coding approaches such as source coding, channel coding which are used to improve information transmitting rate. Network coding is a coding conception first proposed in [1]. Unlike traditional store-and-forward packet delivery technique, network coding implements store, code, and forward approach, where each node stores incoming data packets in its buffer, combines information in different data packets when coding condition appears, then sends coded packets. It is proved in [1] that network coding allows the communication to achieve higher throughput.

Let us consider the benefit of network coding in multicast routing as in Fig. 1. Source node S wants to multicast two data packets P_1 and P_2 to destination node D_1 and D_2. Without using linear network coding, a total of 14 transmissions are needed as in Fig. 1(a). By using linear network coding, source node S can also multicast the same two data packets to D_1 and D_2 as in Fig. 1(b). Firstly, source node S sends two linearly combined packets. Then, if each forwarding node receives two different incoming packets it will recombine those packets with different coding coefficients. Otherwise, it simply forwards the previously received code packet. Destination node

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_17, © Springer-Verlag Berlin Heidelberg 2014

can retrieve data packets P1 and P2 if they receive two coded packets with linear independent coefficients. As we can see in Fig. 1(b), a total of 8 transmission are needed compared with 14 transmissions in the case of without linear network coding.

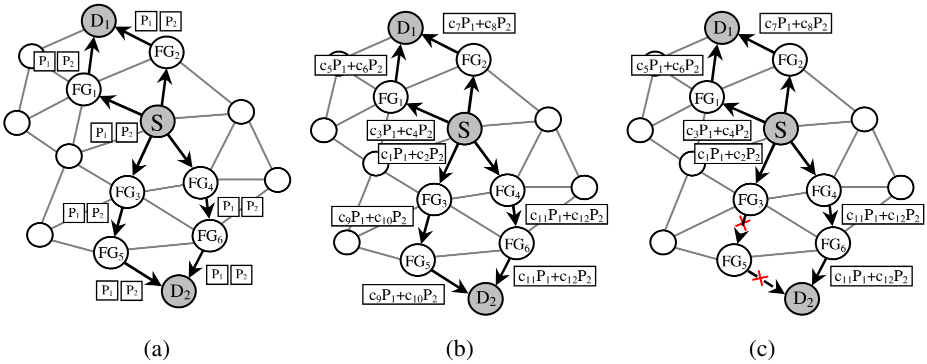


Fig. 1. (a) Multicasting of data packets without network coding, (b) Multicasting of data packets with network coding, and (c) The impact of route break on linear network coding

However, we observe an important fact that in dynamic networks like mobile ad-hoc networks network coding may not increase network throughput or may degrade network performance if multicast destination nodes may not receive enough coded packets to retrieve original packets. Let us illustrate that issue by using Fig. 1(c). Due to node mobility, the wireless link FG_3-FG_5 is broken. Thus, destination node D_2 cannot receive two linearly independent coded packets to recover data packets P_1 and P_2.

Recently, researchers focus on providing efficient multicast communication by using network coding. In [2], a mathematical model is presented to calculate the upper bound value of multicast capacity with network coding. In [3], a network coding based ad hoc multicast protocol, called CodeCast, is proposed. However, CodeCast suffers from decoding failure because of link break due to node mobility as previously mentioned in Fig. 1(c). The authors in [4] use both opportunistic routing and network coding to improve network performance. Since link's ETX metric and Steiner tree are used in [4], it requires a lot of control overheads to maintain network connectivity. In [5], the authors improve the performance of ODMRP by applying network coding. They only apply network coding to Join Request packets. Thus, the coding efficiency is not high. With motivations by all issues in previous works, in this paper we propose a practical Network Coding based Multicast Routing, called NCMR for short, in mobile ad-hoc wireless networks with novel packet coding-and-forwarding decision to improve network throughput while reducing network congestion.

The rest of this paper is organized as follows. In Section 2, we present in detail the architecture of our NCMR including packet format and routing-coding algorithm. In Section 3, we evaluate its performance in different settings of node mobility and the number of multicast member nodes and compare with (i) ODMRP [6], a traditional multicast routing protocol without network coding and (ii) CodeCast [3], a previously proposed network coding based multicast routing protocol. Finally, Section 4 concludes the paper.

2 The Proposed Network Coding Based Multicast Routing

In this section, we present in detail the architecture of our proposed network coding based multicast routing, called NCMR, which includes the format of data packet and step-by-step routing and coding algorithm.

2.1 Packet Format

Our proposed network coding based multicast routing inserts coding header in each coded data packet as in Fig. 2. Multicast ID is the identify number of each multicast group. To perform linear network coding, source node divides data packet stream into batches. Each batch contains M data packets. Batch Size is set to M which can be changed. A data packet p_k is said to be in the batch (*Batch ID*, *Batch Size*) if *Batch ID* $\leq k \leq$ *Batch ID* + *Batch Size*. Code vector has vector size equal to the value in Batch Size. All elements in code vector are random integer values from 1 to 9. Src ID is the ID of source node in multicast group. Sender ID is the ID of node sending the data packet. The list of multicast receivers' IDs is also specified in code header.

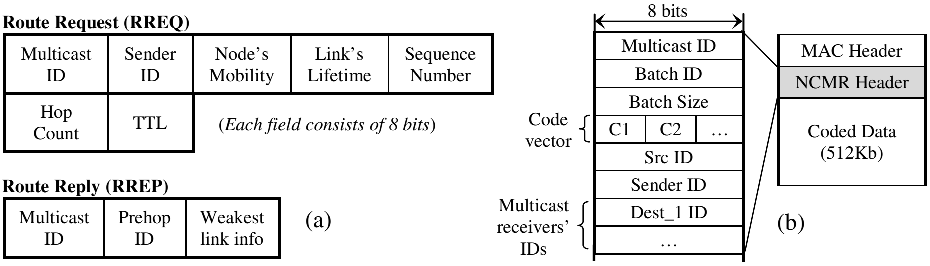


Fig. 2. (a) The format of control overheads and (b) data packet in NCMR

2.2 Network Coding Based Mobile Ad-hoc Wireless Multicast Routing

We use the concept of mesh structure in ODMRP [6] to exploit spatial diversity of forwarding nodes in mesh. We also apply our previous work [7] to enhance the stability of mesh structure. Fig. 3(a) illustrates the basic concept of our proposed NCMR. Fig. 3(b) presents the algorithm to prevent decoding failure at destination nodes.

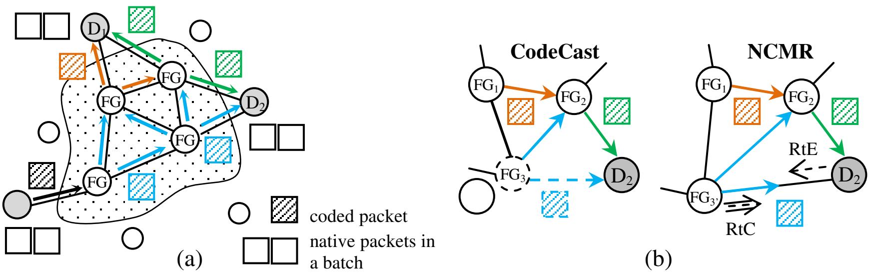


Fig. 3. (a) The basic concept of data packet coding in NCMR, (b) The algorithm to prevent decoding failure at destination nodes

As we can see in Fig. 3(b), NCMR can repair local connection by using Route Expore (RtE) packet and Route Construct (RtC) packet to prevent decoding failure at destination node. RtE and RtC have the same format as RREQ and RREP, respectively, except that their TTLs are set to two hops. Next, we will present step-by-step routing and coding algorithm of NCMR with decoding failure awareness as follows:

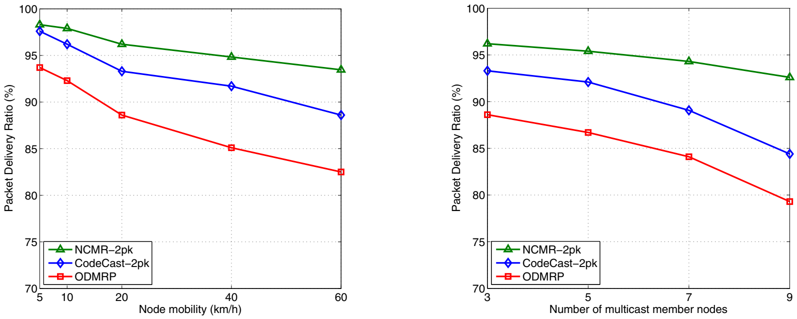
- **Step 1:** Firstly, source node sends Join Request (JREQ) packets with its mobility information throughout the networks. Based on mobility information, intermediate nodes received RREQ will calculate link lifetime which is presented in detail in [7]. After multicast receivers receive JREQ, they will send Join Reply (JREP) to source node via the most stable paths among available paths. Those forwarding nodes on stable paths form routing mesh to deliver data packets from source node to multicast destination nodes.
- **Step 2:** When source node receives JREP packet, it sets route refresh interval based on weakest link information and splits original data packet stream into batches with M packets in each batch. After sending M packets in a batch, source node increases Batch ID of the next batch. All data packets in a batch will be multiplied with random integer coefficients called elements of code vector.
- **Step 3:** Every time forwarder receives coded data packet, it will check Batch ID in that packet to check whether it is from new bath. If yes, it flushes all packets belonging to old batch in its buffer. Otherwise, it checks if that packet is *innovative packet* or not. If an innovative packet is received, it will be stored in forwarder's buffer to be coded with other packets. If not, it will be discarded. After a specific time, forwarding node will combine coded packets together with random coding coefficients.
- **Step 4:** When multicast destination nodes receive coded packet, it performs the same steps as forwarder nodes. However, if the number of received coded packets with the same batch size equals to the number written in Batch Size field of data packet, destination node will decode those coded packet to retrieve native packets and put them to upper layer. If destination node cannot decode the packet due to lack of coded packet, a process of sending RtE and RtC is triggered to create additional path.

3 Performance Evaluation

To evaluate the performance of NCMR, we implement NCMR by using OPNET with the following settings: one multicast group with one source node and varying number of multicast destination nodes; variable number of mobile nodes randomly placed in 1000m×1000m network area; simulation time is 300 s; transmission range is 250 m; 802.11 MAC; data packet size is 512 bytes; data packets are transmitted with constant-based-rate of 20 packets/s; mobile nodes move with Random Waypoint mobility; pause time is randomly selected from 0~10 s; varying maximum speed (5, 10, 20, 40, 60 km/h). Each scenario is simulated 5 times with different node mobility and number of multicast member nodes. The average values are plotted in the graphs.

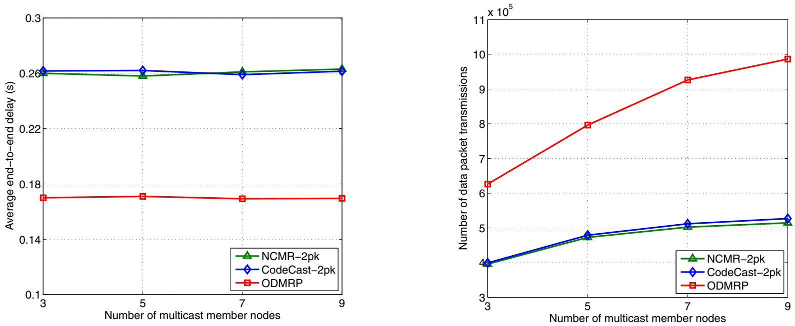
Fig. 4(a) shows the packet delivery ratio (PDR), i.e. the ratio of the number of packet received by multicast destination nodes to those generated by source node, as a function of node mobility. Number of multicast member nodes are 3. As we can see in

Fig. 4(a), the proposed NCMR has higher PDR than CodeCast and ODMRP, especially when node mobility is high because of enhanced stability of mesh structure and an algorithm to fix decoding failures as mentioned above. Fig. 4(b) shows the PDR as a function of the number of multicast member nodes in networks which also refer as the scalability of multicast routing protocols. Node mobility is 40km/h. As we can see in Fig. 4(b), NCMR has significantly higher scalability than CodeCast and ODMRP as the number of multicast member nodes increases. The reason is that NCMR uses linear network coding together with optimal route refresh interval to reduce network congestion. NCMR also has a mechanism to prevent decoding failures for all multicast destination nodes.



(a) (b)

Fig. 4. PDR as functions of node mobility and the number of multicast member nodes



(a) (b)

Fig. 5. Average end-to-end delay and number of duplicated data packet as function of the number of multicast member nodes

Fig. 5(a) shows the drawback of NCMR, the average end-to-end delay. In NCMR, the increase of average end-to-send delay, which also happens in CodeCast, is unavoidable because at each forwarding node it takes time to collect enough appropriate packets to be coded together. However, the gain from the numbers of native packets in coded packet will compensate this drawback. Fig. 5(b) shows the number of duplicate data packet as a function of the number of multicast member nodes with node mobility of 20km/h. As we can see in Fig. 5(b), NCMR generates significantly lower number of data packet transmissions than ODMRP. As the number of multicast member nodes increases, the gap in the number of duplicate data packet of our proposed NCMR and ODMRP is larger. NCMR has moderately lower number of data transmissions because it can reduce the number of forwarding node by using optimal route refresh interval.

4 Conclusion

In this paper, we propose a robust Network Coding based Multicast Routing, named NCMR, in mobile ad-hoc wireless networks. Firstly, a stable mesh topology is established to provide reliable packet delivery from source node to multicast destination nodes in dynamic network topologies as mobile ad-hoc wireless networks. Secondly, we apply linear network coding to forwarders in mesh. By applying linear network coding only to forwarders in mesh topology, we exploit spatial diversity to improve coding/decoding opportunities. Finally, each forwarder checks coding and decoding opportunity carefully instead of blindly code packet together. The simulation results show that the performance of our proposed network coding based multicast routing protocol outperforms with that of CodeCast and ODMRP.

Acknowledgment. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (20120007119) and by National Research Foundation of Korea (NRF) funded by Korean government (MEST) (2012046780).

References

1. Ahlswede, R., Cai, N., Li, S., Yeung, R.W.: Network Information Flow. IEEE Transactions on Information Theory 46(4), 1204–1216 (2000)
2. Wang, Z., Karande, S.S., Sadjadpour, H.R., Garcia-Luna-Aceves, J.J.: On the multicast capacity of wireless ad hoc networks with network coding. IEEE Journal of Communications and Networks 13(5), 525–535 (2011)
3. Park, J.S., Gerla, M., Lun, D., Li, Y., Medard, M.: CodeCast: A Network Coding based Ad hoc Multicast Protocol. IEEE Wireless Communications 13, 76–81 (2006)
4. WenZhong, Y., Chuanhe, H., Wang, B., Zhenyu, Z., Tong, W.: A reliable multicast for MANETs based on opportunistic routing and network coding. In: Proc. of IEEE WCNIS 2010, pp. 540–554 (2010)
5. Yang, T., Gang, W., Qiang, Z., Jian, L., Jun, W.: Improved network coding based on ODMRP protocol in Ad Hoc network. In: Proc. of International Conference on Advanced Computer Control (ICACC), pp. 461–464 (2011)
6. Lee, S.-J., Gerla, M., Chiang, C.-C.: On-Demand Multicast Routing Protocol. In: Proc. of IEEE WCNC 1999, pp. 1298–1302 (1999)
7. Dung, L.T., Ha, S.H., An, B.: A Practical Adaptive Scheme for Enhancing Network Stability in Mobile Ad-Hoc Wireless Networks. In: Park, J.J(J.H.), Arabnia, H.R., Kim, C., Shi, W., Gil, J.-M. (eds.) GPC 2013. LNCS, vol. 7861, pp. 886–892. Springer, Heidelberg (2013)

An Efficient Car Sharing Service in Vehicular Ad Hoc Network

Weicheng Zhao, Deyun Gao, Yajuan Qin, and Wanting Zhu

Beijing Jiaotong University, Beijing 100044, P.R. China

Abstract. With the increasing number of vehicles, traffic congestion and environmental pollution is getting worse. Domestic structures and international institutions pay much attention to the green driving technologies like car sharing service in Vehicular Ad Hoc Network (VANET). To promote fuel efficiency and environmental friendliness, this paper propose an efficient car sharing service based on dual Social Group Architecture (SGA) in VANET. Vehicles generate Drivers Social Group Architecture (DSGA) message and Vehicles Social Group Architecture (VSGA) message to complete car sharing service. We make a basic geometry matching by generating the DSGA messages. Meanwhile, after generating VSGA message, the final matching strategy is processed. The detection in VSGA message is finished by the limited neighbors to decrease the network consume. Our scheme shorten the waiting time for the passengers and avoid traffic jams.

Keywords: car sharing, vehicle ad hoc networks, Social Group Architecture.

1 Introduction

Nowadays, more and more scholars begin to focus on in-vehicle information service [1][2][3]. Car sharing service is the most widely used by public . Car sharing service is used to reduce the use of private cars in order to ease traffic congestion and CO2 emissions. Such services for urban traffic brings many positive economic and environmental effects. Many information services are created to improve the way of his performance.

Traditional distributed services focus on enhancing the utilization of spare seats to save fuel and reduce traffic congestion [4–6].Most of the methods concerns driver and passengers from the point of view, that is, the passenger of car sharing is convenient and fuel-saving in a geometry way.But when the road between the driver and passenger is congested particularly, these methods will face some problems[7]:

1.The vehicles may get idle along the more congested path to pick passengers. A lot time and cost additional fuel will be wasted.

2.Passengers need to wait for very long time, and they can not predict the specific vehicle's arrival time.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_18, © Springer-Verlag Berlin Heidelberg 2014

To solve these problems, in this paper, we propose an efficient car-sharing service. We first divide the two identities, namely, Drivers Social Group Architecture (DSGA) identifier and Vehicles Social Group Architecture (VSGA) identifier.The former consider the relationship between the driver and passengers, the later identifies the vehicle traffic areas.This will be described in detail later.

The rest of the paper is organized as follows.We gives the main framework of our dual-SGA based distributed information service in Section 2. Section 3 introduces how the DSGA message generate and the basic car sharing request matching strategy. Section 4 presents the VSGA identifier based information generation and final matching make. We conclude the contribution in Section 5 finally.

2 System Model

First of all, we carefully designed a carpool matching service system for Vehicle Ad Hoc Network (VANET). VANET contains a large number of vehicles equipped with high-precision sensor nodes.

In this paper, the author propose a new and highly efficient ride sharing service. In car sharing Service, we give two kind of identifier messages: the DSGA based message and VSGA based message. VSGA information represents the traffic congestion area of vehicles. DSGA message reflects the relationship between the members of a set of drivers and passengers between the current location and destination. The DSGA message allows passengers ride in the same car the same destination, which enhance the utilization of the vehicle. And VSGA message helps the car to avoid congestion, thereby reducing the time spent on the road and fuel consumption.

Fig. 1 shows our basic information service system , the neighbor vehicles means the token vehicles' neighbor.There are excursions willingness to carpool

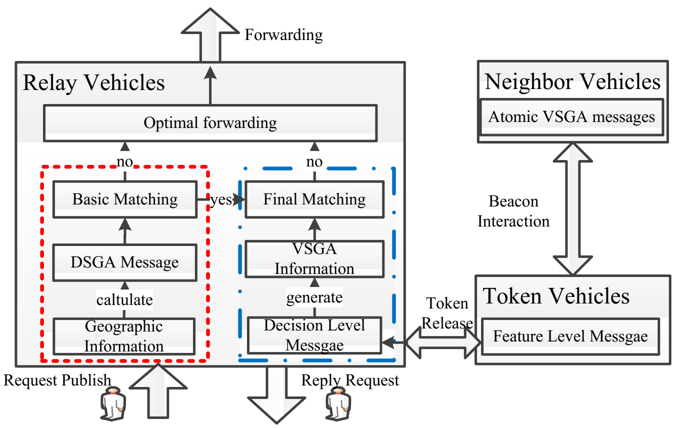


Fig. 1. System framework

passengers will be the first release request, which contains the purpose of their current geographic information and destination geographic information.A generated information will be forwarded by relay vehicles, which are equipped with the device supporting 802.11p protocol.After that, the vehicles begin a DSGA matching step, that is, each vehicle will carry out the basic geometric match procedure.We will continue through the token vehicles forwarding the request proliferation.According VSGA identifier, each vehicle node will monitor and collect the surrounding traffic information, which can accurately and timely reflect the traffic situation around.

3 DSGA Message Generation with Basic Matching Step

DSGA identifier message analyze the drivers and passengers in the carpooling service's role. Among them, carpool service starts with passengers publishing carpooling requests and finish with the final match completion. As shown in figure. 2. passengers release carpool request letter and forwarded by the relay vehicle.If the match is completed, the vehicle will reply the carpool matching requests through the relay vehicles.

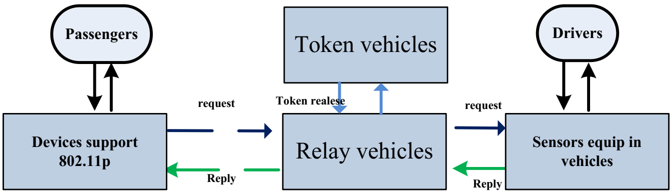


Fig. 2. Information Matching Steps

When the relay vehicle receives the request, DSGA procedure will first do the geographic matching calculation.If the match is successful, the vehicle will do a VSGA identity match step.If the match fails, the request will be forwarded by the relay vehicle.In the vehicle geographic matching calculation, we evaluated the driver and passenger destination correlation. According to figure 3 , $P_L P_D$ reflects the direction vector from present location P_L to the destination P_D of the passengers . Meanwhile , $D_L D_D$ represents the direction vector from present location to the destination of the driver. *theta* reflects the intersection angel between $P_L P_D$ and $D_L D_D$.we have the equation that:

$$\cos\theta = \frac{{d_{P_L D_D}}^2 + {d_{S_L S_D}}^2 - {d_{D_D P_D}}^2}{2 d_{P_L D_D} \cdot d_{P_L P_D}} \tag{1}$$

Where $d_{P_L D_D}, d_{P_L P_D}, d_{D_L D_D}$ separately describe the distance among the current location of the passengers or driver and the destinations of the passengers and driver. The value of $cos\theta$ ranges from 0 to 1, the higher value means the relate direction of passengers and vehicles is more similar.

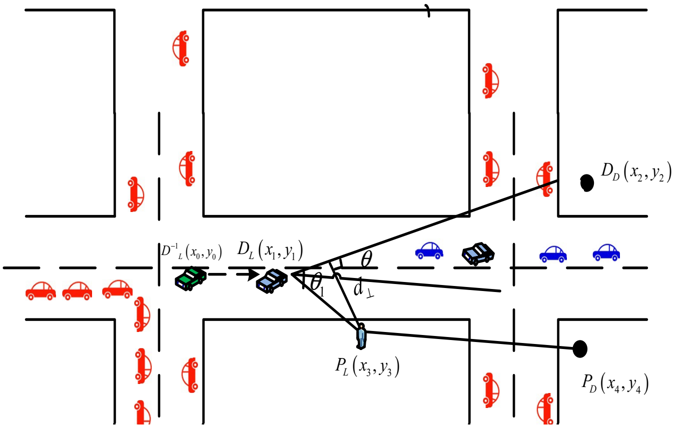


Fig. 3. Basic Geometry Matching

4 VSGA Information Generation with Ultimate Matching Algorithm

VSGA information reflects the whole road traffic congestion information between passengers and vehicles. As shown in Fig. 4. each relay vehicle merge and collect their own traffic information with the help of their neighbor vehicles.For each relay vehicle , we give n tokens is used to control the peripheral congestion message in final match step. The relay vehicle through the beacon package will be token to the neighbor vehicles.With the beacon package neighbor vehicles will collect the traffic information and send to the relay vehicle by the beacon packet.Relay vehicle will do a Fuzzy fusion resulting in regional congestion information. Then, the neighbor vehicle return the token and embrace the message to the other relay vehicle.Finally, the relay vehicle will collect all the regional congestion information along the way and converge to VSGA message.Matching vehicle receives all the relay car VSGA messages, which can reflect the whole road's detailed path information timely and accurately.

Finally the relay vehicles make the decision-level information fusion and generate its own VSGA identifier information. The matching vehicles received all the VSGA identifier information from relay vehicles which reflect the price and traffic conditions.

As is shown in Fig. 5, another key part of the DSGA framework is the VSGA message generation. To improve the matching accuracy at the same time,we can not bring too much bandwidth and computation cost. Therefore, we propose a layered congestion monitoring method to collect congestion information. The relay vehicle first releases a number of tokens to his neighbor vehicles. The neighbor vehicle received tokens will collect the atomic congestion message (contains the accelerate, speed and brake frequency) in its driving region, and then do a fuzzy clustering for the message. By extracting key information, fuzzy clustering method can reduce the redundancy of information and computation.

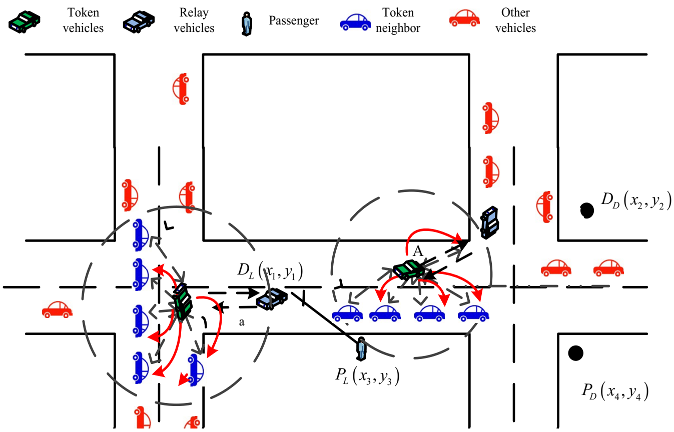


Fig. 4. VSGA Information Matching Structure

All the clustering message will be forwarded to relay the vehicle through the beacon packet.

Every relay vehicle will receive their neighbor vehicles' fuzzy clustering news atomic message results. The number of the results is dominated by the number of tokens. By the decision level fusion, each relay vehicle will generate their own VSGA information. VSGA information reflects the relay vehicle's traffic condition in the congestion area, the accurate is proportional to the number of tokens. All VSGA messages will converge to the matching vehicle. If the geometry matching is successful, the vehicle will do a final carpool matching based on all VSGA message . VSGA information reflects the congestion of the whole road between driver and passenger, which help the vehicle driver determine whether to pick the passenger.

The final VSGA messages matching will be accomplished through the analysis of road congestion. By extracting the all traffic condition between the vehicles and request passengers, we make a final decision about the road congestion level

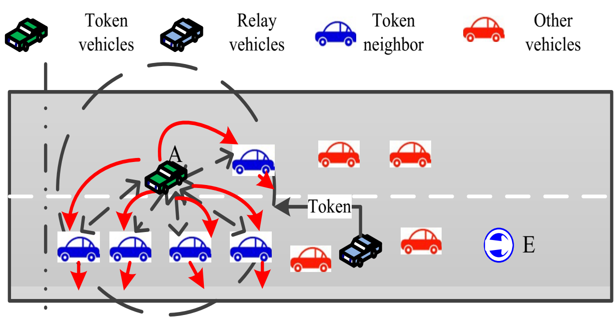


Fig. 5. Traffic congestion happened on two-way road

to decide whether to complete the carpool matching. At the same time, the vehicle calculate the precise time the way to passenger destinations embedded in the reply packet. This scheme greatly improves the matching efficiency and reduce the waiting time of passengers.

5 Conclusions

Finally, the paper closes with the conclusions drawn from those results.A highly efficient distributed car pooling service method is proposed. This shceme enhances the quality and robust of car sharing service which also shortens the the passengers' waiting time and avoids traffic congestion.

References

1. Martin, E., Shaheen, S.A., Lidicker, J.: Impact of carsharing on household vehicle holdings. Transportation Research Record: Journal of the Transportation Research Board 2143(1), 150–158 (2010)
2. Agatz, N.: Optimization for dynamic ride-sharing: A review. European Journal of Operational Research 223(2), 295–303 (2012)
3. Kuntzky, K., Wittke, S., Herrmann, C.: Car and Ride Sharing Concept as a Product Service System – Simulation as a Tool to Reduce Environmental Impacts. The Philosopher's Stone for Sustainability, pp. 381–386. Springer, Heidelberg (2013)
4. Agatz, N., Erera, A.L., Savelsbergh, M.W.P., et al.: Dynamic ride-sharing: A simulation study in metro Atlanta. Procedia-Social and Behavioral Sciences 17, 532–550 (2011)
5. Chung, C.L., Jeng, J.Y., Lee, Z.Y., et al.: Study of Carpool User Behaviors and Route Characteristics in Taiwan. In: Sustainable Transportation Systems@ sPlan, Design, Build, Manage, and Maintain, pp. 356–363. ASCE (2012)
6. Liu, N., Liu, M., Cao, J., Chen, G., Lou, W.: When transportation meets communication: V2p over vanets. In: 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 567–576. IEEE (2010)
7. Sheu, J.-P., Chang, G.-Y., Chen, C.-H.: A distributed taxi hailing protocol in vehicular ad-hoc networks. In: 2010 IEEE 71st Vehicular Technology Conference (VTC 2010-Spring), pp. 1–5. IEEE (2010)

Novel Protocols of Modulation Level Selection in Decode-and-Forward Multinode Cooperative Communication Systems

Kuo-Kuang Wu[1], Min-Kuan Chang[2], Shi-Yong Lee[1], and Yu-Wei Chan[3]

[1] Department of Electrical Engineering, National Chung Hsing University, Taiwan
[2] Graduate Institute of Communication Engineering, National Chung Hsing University, Taiwan
[3] Department of Information Management,
Chung Chou University of Science and Technology, Taiwan
{wkg,minkuanc}@dragon.nchu.edu.tw, sylee@wmc.ee.nchu.edu.tw,
ywchan@dragon.ccut.edu.tw

Abstract. The efficiency of multinode cooperative communication systems is an important issue to solve. To this end, the adaptive modulation (AM) is adopted. However, with the decode-and-forward (DF) protocol, the uncertainty of the decoding behavior of each relay complicates the support of AM. In this work, we first propose two protocols which are the optimal protocol and the suboptimal protocol to maximize the achievable data rate of the current transmission based on the instantaneous channel state information (CSI) of all links and to maximize the long-term achievable data rate, respectively. Finally, simulation results show the advantages of the proposed protocols.

Keywords: Multinode cooperative communication, adaptive modulation, decode-and-forward protocol, relay selection.

1 Introduction

Diversity techniques have been widely adopted to mitigate the channel impairments in a wireless environment. Conventionally, the multiple-input-multiple-output (MIMO) system has been drawn much attention [1] and [2]. However, it is not cost efficient. Thus, cooperative communication is employed to let a group of mobile devices share their own antenna and form a virtual antenna array. The behavior and performance of cooperative communication has been studied extensively [3][4] and its potential benefit to the system has been investigated such as lifetime improvement [5] and coverage extension [6].

Although the benefits of cooperative communication, the reduction in the spectral efficiency needs to be restored. The adaptive modulation (AM) is one of the possible solutions [7]. This concept has been adopted to improve the spectral efficiency in both the relay networks and the cooperative communication. But the works mostly focused on how to provide AM when AF protocol is made use of in the system and none of them considered the same problem in the case of the decode-and-forward (DF)

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_19, © Springer-Verlag Berlin Heidelberg 2014

protocol. The main challenge comes from the uncertainty of the decoding results of relays. In this work, two protocols are proposed to overcome the challenge and to select the most suitable modulation level.

In this work, we propose two protocols to have AM work in the multinode cooperative communication systems with DF protocol, which are the optimal protocol and the suboptimal protocol, respectively. First, the main objective of the optimal protocol is to select the maximal feasible modulation level while the instantaneous SER can be fulfilled. However, the protocol would incur extensive information transmission and reduce the efficiency of the system. Thus, the suboptimal protocol is then proposed, in which the channel distribution information (CDI) of relays is required. Based on these CDI's, the destination chooses the maximal expected feasible modulation level such that the long term SER can be met.

This paper is organized as follows. First, a system model is proposed in Section 2. Then, the proposed protocols and modeling are presented in Section 3. The simulation results are presented in Section 4. The concluding remarks are given in Section 5.

2 System Model

We consider a multinode cooperative communication system which consists of N+2 nodes. In addition to the source (S), and the destination (D), there are N relays ($\mathcal{R}^{(i)}$, i=1,2,⋯,N) deployed in the system. These relays form a relay set $\Omega = \left\{R^{(i)}\right\}$. To simplify the notation in some cases, we use either S or $\mathcal{R}^{(0)}$ to refer to the source in this work. DF protocol is employed at each relay [4], which helps the information forwarding from S to D. A relay only helps forward the data if and only if the received symbol can be correctly decoded. At D, it will make use of maximum ratio combining (MRC) to combine all received signals from S and relays. In this work, the weights of MRC are set according to [4]. Furthermore, $d_{sd}, d_{sr}^{(i)}$, and $d_{rd}^{(i)}$ are the propagation distance of the links S – D, $\mathcal{S}$–$\mathcal{R}^{(i)}$, and $\mathcal{R}^{(i)}$–$\mathcal{D}$, respectively. $\theta^{(i)}$ is the angle between $\mathcal{S}$–$\mathcal{D}$ and $\mathcal{S}$–$\mathcal{R}^{(i)}$.

2.1 Channel Model

The channel effect consists of two factors, namely path loss and fading. The path loss in this work follows the log-distance path loss model [8]. The fading process adopted in this work follows the FSMC of K states in [9]. Each FSMC state defines a range of received SNR rather than the actual fading gain [9]. Since the path loss is invariant, the aggregate effect of the path loss and the fading process of each link still follow the FSMC of K states. However, now the average received SNR of this FSMC is that of the original FSMC scaled by the path loss effect. To facilitate the derivations to come, we denote $S_r[i] \in \{\mathbf{s}_{r,1}^{(i)}, \mathbf{s}_{r,2}^{(i)}, \cdots, \mathbf{s}_{r,K}^{(i)}\}$ be the FSMC state of $\mathcal{S}$–$\mathcal{R}^{(i)}$ link for i=1,2,⋯,N and $S_d[i] \in \{\mathbf{s}_{d,1}^{(i)}, \mathbf{s}_{d,2}^{(i)}, \cdots, \mathbf{s}_{d,K}^{(i)}\}$ be the FSMC state of $\mathcal{R}^{(i)}$–$\mathcal{D}$ for i=0,1,⋯,N. For simplicity, we further let $S_\tau[i]_1$ and $S_\tau[i]_2$ refer to the FSMC states of two consecutive time slots. The steady state probabilities and the transition probabilities of all links are denoted by $\pi_{S_\tau[i]}$ and $P_{S_\tau[i]_1, S_\tau[i]_2}$, where $\tau \in \{r,d\}$, respectively.

2.2 Adaptive Modulation Scheme

AM scheme is adopted to improve the spectral efficiency. In this work, $\mathcal{D}$ is responsible for choosing the suitable modulation level based on the channel-related information of each relay. In this AM system, $m \in \{1,2,3,\cdots,M\}$ is the modulation level for 2^m-ary phase shift keying (2^m-PSK) modulation with maximum allowable order M, P_s^{req} represents the required symbol error rate (SER), and the associated SNR thresholds is given by

$$\Gamma_m^{TH} = \frac{1}{2}\left(\frac{Q^{-1}\left(P_s^{req}/2\right)}{\sin\pi/2^m}\right)^2,$$

for all m, which is based on the SER of 2^m-PSK [10] and $Q^{-1}(\cdot)$ is the inverse Q-function.

According to the above description, we have the modulation level be m if and only if

$$\Gamma_m^{TH} \le \gamma_{eq} < \Gamma_{m+1}^{TH}, \tag{1}$$

where γ_{eq} is the equivalent total SNR of the multinode cooperative communication system when MRC is utilized at $\mathcal{D}$. Moreover, we say the system is cut-off if no modulation level can be supported, namely, $\gamma_{eq} < \Gamma_1^{TH}$. Here we define two kinds of relay sets: One is the participating relay set Ω_p, $\Omega_p \subseteq \Omega$, and the other is the decoding relay set Ω_d, $\Omega_d \subseteq \Omega$. The participating relay set contains those relays that are selected to participate the current transmission by $\mathcal{D}$.

Next, let z be the combined signal after MRC based on the signals from $\mathcal{S}$ and $\mathcal{R}^{(i)} \in \Omega_p$.

$$z = \frac{h_d^{(0)^*}}{N_0} \cdot y^{(0)} + \sum_{\{i|R^{(i)} \in \Omega_p\}} \frac{h_d^{(0)^*}}{N_0} \cdot y^{(i)}.$$

Then, the corresponding equivalent total SNR can be written as

$$\gamma_{eq} = \gamma_d^{(0)} + \sum_{\{i|\mathcal{R}^{(i)} \in \Omega_p\}} \gamma_d^{(i)}, \tag{2}$$

3 Proposed Protocols

3.1 The Optimal Protocol

The optimal protocol aims at achieving the highest achievable data rate as possible while the SER requirement at $\mathcal{D}$ can be fulfilled as well. In this protocol, we assume that the CSI information of all links is available to $\mathcal{D}$. To develop the optimal protocol, we first calculate the SER's given that the participating relay set is Ω_p^u and the modulation level is $\mathbf{m}=\Phi\left(\Lambda_u^T\gamma_d\right)$. Let $P_{s|\Omega_p^u,\gamma_r,\gamma_d}$ be the corresponding SER conditioned on γ_r and γ_d. To find $P_{s|\Omega_p^u,\gamma_r,\gamma_d}$, all possible decoding scenarios of $2^{\boldsymbol{m}}$-PSK symbol at relays should be taken into account. Then, $P_{s|\Omega_p^u,\gamma_r,\gamma_d}$ can be found by

$$\begin{aligned} P_{s|\Omega_p^u,\gamma_r,\gamma_d} &= \sum_{u'=1}^{2^N} \Pr\left\{\text{Symbol error}|\gamma_{eq}=\tilde{\boldsymbol{\Lambda}}_{u'}^T\gamma_{\boldsymbol{d}}\right\}\cdot \mathrm{P} \\ &= \sum_{u'=1}^{2^N} 2Q\left(\sqrt{2\tilde{\boldsymbol{\Lambda}}_{u'}^T\gamma_{\boldsymbol{d}}}\sin\frac{\pi}{2^{\mathbf{m}}}\right)\cdot P_{\Omega_{\mathbf{m}}^{u'}|\Omega_p^u,\gamma_r}, \end{aligned}$$

where $P_{\Omega_{\mathbf{m}}^{u'}|\Omega_p^u,\gamma_r}$ is given by

$$P_{\Omega_{\mathbf{m}}^{u'}|\Omega_p^u,\gamma_r} = \prod_{i=1}^{N}\left[\lambda[u',i]+(-1)^{\lambda[u',i]}2Q\left(\sqrt{2\gamma_r^{(i)}}\sin\frac{\pi}{2^{\mathbf{m}}}\right)\right]$$

After computing $P_{s|\Omega_p^u}$ under all possible participating relay sets, Ω_p^u for all u, the optimal participating relay set can be determined according to the following rule.

$$\Omega_p^{u^*}=\operatorname{argmax}_{\Omega_p^u}\left\{\Phi\left(\Lambda_u^T\gamma_d\right)\right\} \quad \text{s.t. } \mathrm{P}_{\mathrm{s—\Omega_p^u,\gamma_r,\gamma_d}}\le\mathrm{P_s^{req}},$$

This rule maximizes the current achievable data rate and can satisfy the SER requirement in the long run given γ_r and γ_d.

3.2 The Suboptimal Protocol

The prerequisite condition for the success of the optimal protocol is the availability of CSI information of all links at $\mathcal{D}$ for each transmission. However, to make this information available would incur the extensive information dispatch between a relay and $\mathcal{D}$. This would in turn decrease the efficiency of the system. To avoid this, in this protocol, the CDI information of all links are utilized instead, which is why this protocol is called the suboptimal protocol. The CDI information can be acquired and dispatched in the beginning. Once the CDI information is obtained by $\mathcal{D}$, such information dissemination is no longer needed.

Similar to the optimal one, we would like to find a participating relay set that can provide better achievable data rate and can fulfill the SER requirement. Instead of finding the instantaneous modulation level, $\mathbf{m}=\Phi\left(\Lambda_u^T\gamma_d\right)$, and the instantaneous SER, $P_{s|\Omega_p^u,\gamma_r,\gamma_d}$, $E\left\{\Phi\left(\Lambda_u^T\gamma_d\right)|\Omega_p^u\right\}$ and $P_{s|\Omega_p^u}=E\left\{P_{s|\Omega_p^u,\gamma_r,\gamma_d}\right\}$ are taken into account in the suboptimal protocol. The rule of selecting the participating relay set in the suboptimal protocol becomes

$$\Omega_p^{u*}=\mathrm{argmax}_{\Omega_p^u}\left\{E\left\{\Phi\left(\Lambda_u^T\gamma_d\right)\middle|\Omega_p^u\right\}\right\} \quad \text{s.t. } P_{s|\Omega_p^u}\le P_s^{req},$$

where

$$E\left\{\Phi\left(\Lambda_u^T\gamma_d\right)\middle|\Omega_p^u\right\}=\int_{\forall\gamma_d}\Phi\left(\Lambda_u^T\gamma_d\right)f_{\Gamma_d}\left(\gamma_d\right)d\gamma_d,$$

$f_{\Gamma_d}\left(\gamma_d\right)$ is the joint PDF of γ_d and $P_{s|\Omega_p^u}$ is given by

$$\begin{aligned}
P_{s|\Omega_p^u} &= \mathrm{E}\left\{P_{s|\Omega_p^u,\gamma_r,\gamma_d}\right\}\\
&= \int_{\forall\gamma_r}\int_{\forall\gamma_d} P_{s|\Omega_p^u,\gamma_r,\gamma_d}\cdot f_{\boldsymbol{\Gamma_r}}(\boldsymbol{\gamma_r})\cdot f_{\boldsymbol{\Gamma_d}}(\boldsymbol{\gamma_d})d\boldsymbol{\gamma_d}d\boldsymbol{\gamma_r}\\
&= \int_{\forall\gamma_d}\sum_{u'=1}^{2^N}2Q\left(\sqrt{2\tilde{\boldsymbol{\Lambda}}_{u'}^T\boldsymbol{\gamma_d}}\sin\frac{\pi}{2^{\mathrm{m}}}\right)\cdot P_{\Omega_{\mathrm{m}}^{u'}|\Omega_p^u}\cdot f_{\boldsymbol{\Gamma_d}}(\boldsymbol{\gamma_d})d\boldsymbol{\gamma_d}.
\end{aligned}$$

4 Simulation Results

In the simulation, we consider the case of that the relays are spaced equally between the $\mathcal{S}$–$\mathcal{D}$ link, i.e., $d_{sr}^{(i)}=i\cdot d_{sd}/(N+1)$ and $\Theta^{(i)}=0$ for simplicity. In such a setting, $\mathcal{R}^{(1)}$ is the nearest relay to $\mathcal{S}$ and $\mathcal{R}^{(N)}$ is the nearest relay to $\mathcal{D}$. Also for convenience, the statistical property of all links are identical, that is $\gamma_0^{(i)}=\tilde{\gamma}_0^{(i)}=\gamma_0$ for all i. The rest of the experimental parameters are: $d_{sd}=50$ m, $\alpha=2$, $d_{min}=10$ m, $P_s^{req}=10^{-3}$, $K=11$, and M=8. The protocols to be compared with the cases, including the ideal case (Ideal), the simplest protocol (Simp.), the SNR-based protocol, the multi-hop case (MH), and the direct link case (DL).

From Fig. 1, we compare the achievable data rate of the proposed two protocols with N=3 and N=5 when there is no total power constraint and has total power constraint. First, in the case of no total power constraint, $\mathcal{S}$ and relays in the system have the same transmission power. As can be clearly seen, the achievable data rate is improved as N increases for both protocols in this case. Such improvement is due to the increased power and the diversity gain. Next, when the total power constraint is imposed, the proposed suboptimal protocol not only has better achievable data rate but also extends the system lifetime as N increases from 3 to 5.

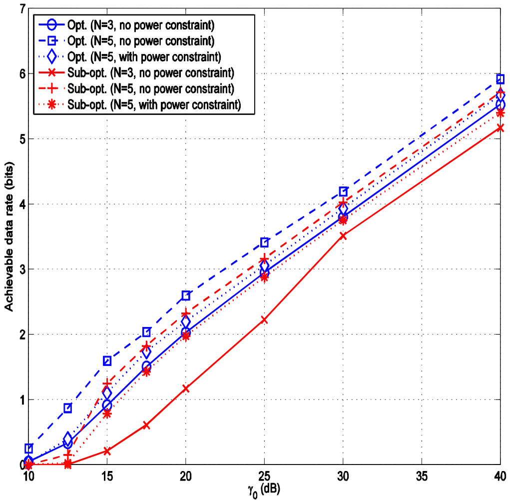


Fig. 1. The comparison of the achievable data rate with and without total power constraint (PC). The number of relays are 3 and 5.

References

1. Dighe, P.A., Mallik, R.K., Jamuar, S.S.: Analysis of Transmit-Receive Diversity in Rayleigh Fading. IEEE Transactions on Communications 51, 694–703 (2003)
2. Kang, M., Alouini, M.-S.: Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems. IEEE Journal on Selected Areas in Communications 21, 418–426 (2003)
3. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative Diversity in Wireless Networks: Efficient protocols and Outage Behavior. IEEE Transactions on Information Theory 50, 3062–3079 (2003)
4. Su, W., Sadek, A.K., Liu, K.J.R.: SER Performance Analysis and Optimum Power Allocation for Decode-and-Forward Cooperation Protocol in Wireless Networks. In: Proceedings of IEEE WCNC 2005 (March 2005)
5. Han, Z., Poor, H.V.: Lifetime Improvement of Wireless Sensor Networks by Collaborative Beamforming and Cooperative Transmission. In: Proceedings of IEEE ICC 2007 (June 2007)
6. Sadek, A.K., Han, Z., Liu, K.J.R.: An efficient cooperation protocol to extend coverage area in cellular Networks. In: Proceedings of IEEE WCNC 2006 (March 2006)
7. Qiu, X., Chawla, K.: On the performance of adaptive modulation in cellular systems. IEEE Transactions on Communications 47, 884–895 (1999)
8. Rappaport, T.S.: Wireless Communication Priciples and Practice. Prentice-Hall (1996)
9. Zhang, Q., Kassam, S.A.: Finite-state Markov model for Rayleigh fading channels. IEEE Transactions on Communications 46(11), 1688–1692 (1999)
10. Proakis, J.G.: Digital Communication. McGraw-Hill (2000)

Fair Spectrum Allocation with Reducing Spectrum Handoff in Cognitive Radio Sensor Networks*

Sang-Seon Byun[1] and Joon-Min Gil[2,**]

[1] Department of Computer Engineering,
Catholic University of Pusan,
Pusan, Republic of Korea
ssbyun@cup.ac.kr
[2] School of Information Technology Engineering,
Catholic University of Daegu,
Gyeongsan, Gyeongbuk, Republic of Korea
jmgil@cu.ac.kr

Abstract. This paper considers the problem of centralized spectrum allocations in wireless cognitive radio sensor networks towards the following goals: (1) maximizing fairness reflecting the priority among sensor data, and (2) avoiding unnecessary spectrum handoff. We cast this problem into a multiobjective mixed integer non-convex nonlinear programming. To tackle the intractability, we first convexify the original problem using arithmetic-geometric mean approximation, and then deploy weighted Chebyshev norm-based scalarization method in order to collapse the multiobjective problem into a single objective one. Finally, we apply simple rounding method in order to obtain approximate integer solutions. A noticeable advantage of the proposed approach is that it enables to find fairly good approximated integer solutions within reasonable computation time.

1 Introduction

The demand of radio frequency spectra is rapidly growing with the current static spectrum allocation. Dynamic spectrum allocation may mitigate this problem by opening assigned, but sparsely used, spectrum resources to secondary users [1]. Such dynamic spectrum access schemes can be considered in wireless sensor networks (WSNs) as well. One of the primary objectives of WSNs is to transmit monitored results timely and concurrently, without using large amount of network resources. For instance, in a WSN for real-time surveillance system, the transmissions of video or image data captured by the sensors requires high bandwidth and multiple spectra [2]. Subsequently, the following major principles can be considered:

---

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A4A01015777).
** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_20, © Springer-Verlag Berlin Heidelberg 2014

- Fair allocation of idle spectrum bands

  Scarce spectrum resources should be allocated as fairly as possible with prioritizing the transmissions if necessary. This can be achieved by maximizing *proportional fairness* with demanding weights [3].
- Avoiding unnecessary spectrum handoff

  The other primary objective of WSNs is long time functionality. However, dynamic spectrum access leads to supplemental energy consumption at each sensor due to *spectrum handoff*. It has been measured that frequency or timing synchronization consumes a certain amount of power [4]. Hence unnecessary spectrum handoff should be eliminated [5, 6].
- Centralized spectrum allocation

  If distributed scheme is deployed to WSN, all sensors should have a fully operating cognitive radio. However it is not feasible to implement full features of cognitive radio in such power-limited sensor nodes. Thus, in a moderate size of WSN, where sensors are not distributed widely[1], the centralized scheme is preferred to the distributed one. In this paper, we consider the problem of the centralized spectrum allocations in a WSN of moderate size, i.e., all sensors are located within a cell or segment boundary.

We locate a dedicated coordinator that allocates each sensor spectrum resources: (1) *as fairly as possible reflecting the priority among sensors*, and (2) *avoiding unnecessary spectrum handoff*. We formulate this problem into a multiobjective – more clearly, bicriteria or biobjective – mixed integer nonlinear non-convex programming that is, however, known as intractable without any aid of modification or approximation.

Our approach to tackle this problem is summarized as follows. First, we convert the original formulation into a quasi-equivalent form that is convex through *arithmetic-geometric mean approximation* and logarithmic change of decision variables. Then we relax the integer constraint, so called, *NLP-relaxation*, and collapse the multiobjective optimization problem (MOP) into a single objective one using *scalarization* based on *weighted Chebyshev norm*[2] *problem* by which we can maintain the convexity and achieve NLP-relaxed solutions that satisfies weak Pareto optimality. Finally, we perform simple rounding algorithm on the NLP-relaxed solutions in a sequence of steps in order to obtain approximated integer solutions.

2 Problem Formulations

We formulate our centralized spectrum allocation problem as an MO-MINLP. Prior to describing the formulation, the following assumptions are made:

1. Available spectrum resources are expressed as a number of spectrum units. E.g., subcarrier in OFDM or subband in multiband ultra wide band (UWB).
2. Each sensor can transmit over noncontiguous frequency bands concurrently [7].

[1] For example, a healthcare system in an intensive care unit.

[2] Also called as *supremum* or *infinity norm* [9].

3. Similar to [8], each receiver sensor declares its desired signal to interference noise ratio (SINR). We consider the spectrum handoff at only a transmitter.

Based on the above assumptions, the MO-MINLP can be described as follows:

Parameters

- V: Set of sensors that request spectrum units for their transmissions.
- S: Set of idle spectrum units.
- L_{is} (binary): It means that, currently, sensor i is synchronized with the spectrum unit s.
- w_i: Weight given in accordance with the priority of sensor i. The priority reflects the demand of spectrum units.
- $P_i^{\max}$: Maximal transmission power available at sensor i.
- $INSR^{\max}$: 1/ $SINR^{\min}$ where $SINR^{\min}$ is the minimal SINR that corresponds to strict QoS constraint.
- G^s_{ij}: Channel gain between sensor i and j over spectrum unit s.

Decision Variables

- x_{is}(binary): It indicates that sensor i occupies spectrum unit s.
- p_{is}: Transmission power for sensor i in spectrum unit s.
- $INSR_{is}$: Reciprocal of SINR of sensor i in spectrum unit s and defined as

$$INSR_{is} = \frac{\sum_{j \in V \setminus \{i\}} x_{js} p_{js} G^s_{jk} + \sigma^2}{p_{is} G^s_{ik}} \tag{1}$$

where σ^2 is noise power.

Objectives

$$\text{Maximize} f_1 = \sum_{i \in V} w_i \ln\left(\sum_{s \in S} x_{is}\right) \tag{2}$$

$$\text{Maximize} f_2 = \sum_{s \in S} \sum_{i \in V} L_{is} x_{is} \tag{3}$$

s.t.

$$c_1 : x_{is} INSR_{is} \le INSR_i^{\max} \text{ for all } i \in V \text{ and } s \in S, \tag{4}$$

$$c_2 : \sum_{s \in S} x_{is} p_{is} \le P_i^{\max} \text{ for all } i \in V, \tag{5}$$

By maximizing f_1, we can achieve maximal proportional fairness in terms of the number of spectrum units allocated to each sensor, and reflect the priority of each sensor simultaneously [3]. By maximizing f_2, we can let each sensor keep holding the spectrum units used in the previous transmission epoch. The constraint c_1 indicates that there is a corresponding SINR threshold that determines whether or not a packet transmission is successful [10]. The constraint c_2 indicates that no sensor can use more transmission power than its maximal available transmission power. However, the problem is intractable due to its non-conveixity and discrete variables. For this reason, we manipulate the objective functions and apply a series of approximations.

Henceforth, we denote the single objective optimization problem with only the objective function f_1 and the constraints as $\Omega(f_1)$, and the one with only f_2 and the constraints as $\Omega(f_2)$, respectively. In addition, we denote the problem with both the objectives and the constraints as $\Omega(f_1, f_2)$.

3 Tackling the Intractability

We begin the convexification process with relaxing the integer constraint, that is, we drop (6).

3.1 Convexification of Ω(f1) and Ω(f2)

Convexification of Ω(f1)
We convert the original function f_1 into an equivalent log-sum-exp function form – that is proved to be convex [8] – by defining $\beta_{is} = \ln(x_{is})$ and $\gamma_{is} = \ln(p_{is})$ for all $i \in V$ and $s \in S$ where $-\infty \le \beta_{is} \le 0$ and $-\infty \le \gamma_{is} \le \ln(P_i^{\max})$. Then $\Omega(f_1)$ is reformulated as

Objective

$$\text{Maximize } f_1^* = \sum_{i \in V} w_i \ln\left(\sum_{s \in S} e^{\beta_{is}}\right) \tag{6}$$

s.t.

$$Cv(c_1) : \ln\left(\sum_{j \in V \setminus \{i\}} \begin{pmatrix} e^{(\beta_{is}+\beta_{js}+\gamma_{js}-\gamma_{is})} G_{jk}^s \left(G_{ik}^s\right)^{-1} \left(INSR_i^{\max}\right)^{-1} \\ +e^{(\beta_{is}-\gamma_{is})} \sigma^2 \left(G_{ik}^s\right)^{-1} \left(INSR_i^{\max}\right)^{-1} \end{pmatrix}\right) \le 0 \tag{7}$$

for all $i \in V$ and $s \in S$,

$$Cv(c_2) : \ln\left(\sum_{s \in S} e^{(\beta_{is}+\gamma_{is})} \left(P_i^{\max}\right)^{-1}\right) \le 0 \text{ for all } i \in V. \tag{8}$$

Nonetheless the reformulated one is not a convex optimization. Therefore, we deploy arithmetic-geometric mean approximation [11]. Then the objective is condensed as

$$\begin{aligned} &\text{Maximize } Cv(f_1^*) = \sum_{i \in V} w_i \ln\left(\prod_{s \in S} \left(\frac{e^{\beta_{is}}}{\alpha_{is}}\right)^{\alpha_{is}}\right) \\ &= \sum_{i \in V} w_i \sum_{s \in S} \alpha_{is} \left(\beta_{is} - \ln \alpha_{is}\right) \end{aligned} \tag{9}$$

where

$$\alpha_{is} = \frac{e^{\beta_{is}}}{\sum_{s \in S} e^{\beta_{is}}}, \text{ for all } i \in V \text{ and } s \in S. \tag{10}$$

Then the above optimization problem can be solved by the *condensation algorithm* given in [11].

Convexification of $\Omega(f2)$

As done in the convexification of $\Omega(f_1)$, we let β_{is} = ln(x_{is}) and γ_{is}= ln(p_{is}) for all $i \in V$ and $s \in S$ where $-\infty \le \beta_{is} \le 0$ and $-\infty \le \gamma_{is} \le \ln(P_i^{max})$. Then f_2 is converted into a convex form:

$$\text{Minimize } Cv(f_2) = \sum_{s \in S} \sum_{i \in V} L_{is} e^{-\beta_{is}}. \tag{11}$$

3.2 Scalarization of the Multiobjective Optimization Problem

For solving MOPs, one of the most wide spread approach is *scalarization.* In this paper, we are interested in maintaining the convexity of the scalarized problem as well as guaranteeing weak Pareto optimality at least. To this end, we deploy a scalarization method based on *weighted Chebyshev norm problem* [9].

If we denote $Cv(f^*_1)$ and $Cv(f_2)$ as g_1 and g_2 respectively, the problem $\Omega(f_1, f_2)$ is converted into a single objective optimization problem by Chebyshev norm problem:
Objective.

$$\text{Minimize } z, \tag{12}$$

s.t.

$$(8), (9)$$

$$c_4 : \delta_1 \times \left(\frac{-\sum_{i \in V} w_i \sum_{s \in S} \alpha_{is} \left(\beta_{is} - \ln \alpha_{is}\right) - g_1^*}{g_{1w} - g_1^*} \right) \le z, \tag{13}$$

$$c_5 : \delta_2 \times \left(\frac{\ln\left(\sum_{s \in S} \sum_{i \in V} L_{is} e^{-\beta_{is}} \right) - g_2^*}{g_{2w} - g_2^*} \right) \le z \tag{14}$$

where g^*_1 is the optimal value when only $Cv(f^*_1)$ is minimized, and g^*_2 is the optimal value when only $Cv(f_2)$ is minimized. Similarly, g_{1w} is the value of $Cv(f^*_1)$ when only $Cv(f_2)$ is minimized, and g_{2w} is the value of $Cv(f_2)$ only $Cv(f^*_1)$ is minimized. We apply α_{is} obtained by solving $\Omega(f_1)$ to c_4. Therefore, prior to solving $\Omega(f_1, f_2)$, we should solve $\Omega(f_1)$ and $\Omega(f_2)$ respectively.

3.3 Rounding Algorithm

We apply a simple rounding method in [11] for finding approximated integer solutions of $\Omega(f_1, f_2)$, $\Omega(f_1)$ and $\Omega(f_2)$ as follows:

Algorithm 1. Rounding algorithm

Step 1.	*best_sol* := 0;
	old_sol := 0;
	$\boldsymbol{\beta} := \{\beta_{is}: \forall i \in V \text{ and } \forall s \in S\}$;
	$\boldsymbol{\gamma} := \{\gamma_{is}: \forall i \in V \text{ and } \forall s \in S\}$;

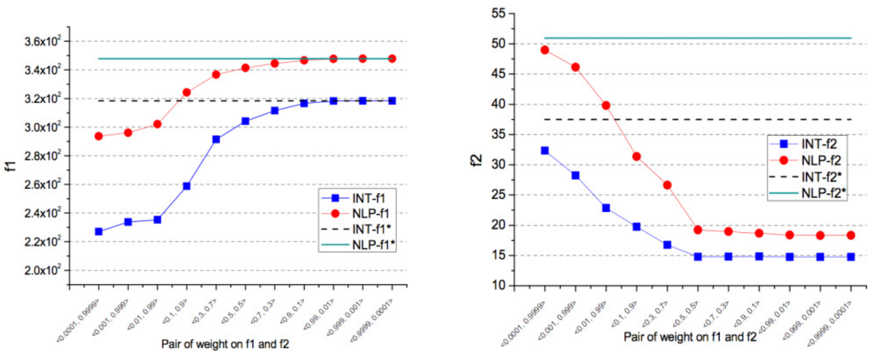
Step 2.	Solve NLP relaxed problem with the solution vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$;

Step 3.	**for each** $\beta_{is} \in \boldsymbol{\beta}$ **begin**
	  if $(\lvert\exp[\beta_{is}] - 1.0\rvert \le \xi)$ x_{is} := 1;
	  else x_{is} := 0;

Step 4. Check the feasibility of the original problem with the rounded solution.
  if not feasible **begin**
    $\ell := \ell - \zeta$; **goto Step 3**;
  end
Step 5. Compute integer solution *int_sol* of each objective function:
  (1) For $\Omega(f_1)$, *int_sol* := $f_1 = \sum_{i \in V} w_i \ln\left(\sum_{s \in S} x_{is}\right)$;
  (2) For $\Omega(f_2)$, *int_sol* := $f_2 = \sum_{s \in S}\sum_{i \in V} L_{is} x_{is}$;
  (3) For $\Omega(f_1, f_2)$, *int_sol* := $\delta_1 f_{1n} + \delta_2 f_{2n}$
  where f_{1n} and f_{2n} are normalized f_1 and f_2, respectively.
Step 6. **if** (*int_sol* > *best_sol*) *best_sol* := *int_sol*;
Step 7. **for each** $\beta_{is} \in \beta$ **begin**
  if ($x_{is} = 1$) $\beta := \beta - \{\beta_{is}\}$; // if $x_{is} = 1$, fix β_{is} as a parameter for next rounding iteration.
  End
Step 8. **if** ($|old\_sol - int\_sol| < \varepsilon$) **Terminate the algorithm**;
  else *old_sol* := *int_sol*; **goto Step 2**;

We observe that the rounding algorithm works well since all the objective functions and constraints involve exponential function[3].

4 Numerical Experiments



(a) Evaluation of f_1 (b) Evaluation of f_2

Fig. 1. Evaluation of f_1 and f_2 achieved by the scalarization method. The label of X-axis on each graph corresponds to the pair of weight on each objective function, $<\delta_1, \delta_2>$. In the legend, 'INT-f1' and 'INT-f2' are associated with the integer optimal values of the corresponding objective functions determined by the scalarization method and rounding algorithm; 'NLP-f1' and 'NLP-f2' correspond to the NLP-relaxed optimal values of the corresponding objective functions obtained after the scalarization. In addition, the items with '*' in the legend correspond to the single objective solutions.

[3] There is a tendency of producing the variables bisected into very small or large values in order to seek the maximum (or minimum) results.

For the experiments, we consider a sensor field of 100m × 100m rectangular area where 15 sensors are uniformly distributed and 60 spectrum units are available for the allocation. We generate each sensor's transmission target randomly. We use the channel gain modeled as $G_{ij} = K_0 \cdot 10^{\delta_{ij}/10} \cdot (d_{ij})^{-v}$, where δ_{ij} is random Gaussian variables with zero mean and standard deviation equal to 6dB, $K_0 = 10^3$ that captures system and transmission effects such as antenna gain, carrier frequency, etc., and d_{ij} is the distance between sensor i and j, and υ is the power falloff factor. We let $\upsilon = 3$.

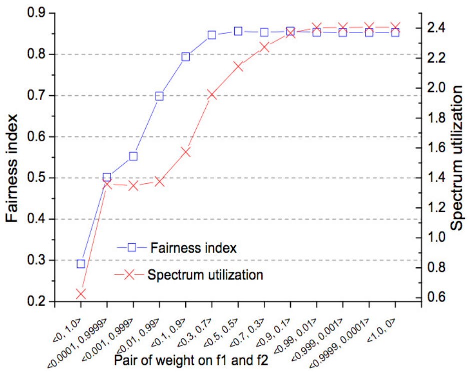


Fig. 2. Fairness index and spectrum utilization according to the weights

We evaluate the multiobjetive solutions determined by the scalarization method by comparing with each single objective solution. Not only the integer solutions but also the NLP relaxed solutions are compared with varying the weight on each objective function, i.e., δ_1 and δ_2. The results are plotted in Fig. 1; we perform this evaluation with 60 different random channels, and plot the average values over those samples. The remarkable results are: (i) both the NLP-relaxed solutions and integer solutions of f_1 and f_2 in $\Omega(f_1, f_2)$ are apparently proportional to their respective weights; (ii) The solutions determined by the scalarization method are strictly bounded to each of the single objective solutions; (iii) The integer solutions are very close to the NLP relaxed solutions with the factor of less than 2 for both the objective functions.

Next we measure how fairly the spectrum resources are allocated. To show this, we compute *fairness index* using (15), and plot it on Fig. 2. We also plot spectrum utilization in order to show how well the spectrum resources are utilized. We plot the average results obtained over 60 different random channels.

$$FI = \left(\sum_{i \in V} \frac{\sum_{s \in S} x_{is}}{w_i} \right)^2 \Bigg/ \left[n \sum_{i \in V} \left(\frac{\sum_{s \in S} x_{is}}{w_i} \right)^2 \right] \tag{15}$$

Fig. 2 shows that, as the weight on f_1 increases, also the fairness index increases generally. It is observed that the highest fairness index is yielded on <0.5, 0.5>, and as f_1 gets more weight, it decreases slightly, and which conforms to the attribute of the proportional fairness; if there are unallocated resources, maximal proportional fairness is achieved as allocating them despite of suffering the deterioration in the fairness index. On the other hand, maximal max-min (or min-max) or strict fairness does not endure such deterioration.

5 Conclusions

This paper deals with the problem of spectrum allocation in resource-constrained wireless sensor networks with the following goals: (1) *maximizing fairness with reflecting the priority among sensor data*, and (2) *avoiding unnecessary spectrum handoff*. Therefore the problem has been formulated as an optimization with two different objective functions: multiobjective optimization. We deploy the scalarization method based on Chebyshev norm problem in order to maintain the convexity of the objective functions and constraints. Prior to applying the scalarization, the original objective functions are convexified by the arithmetic-geometric mean approximation. Also, all the constraints are transformed into log-sum-exp function form that is strictly convex. Furthermore, in order to find the good approximate integer solutions, a simple rounding algorithm is used. By the discreet adjustment of the weight on each objective function, the proposed algorithm performs well in achieving the balanced multiobjective solutions. Finally, it is illustrated that, given the weight of each sensor, the algorithm allocates spectrum units fairly as well as yielding high spectrum utilizations when f_1 has relatively higher weight than f_1.

References

1. Zhao, Q., Sadler, B.: A survey of Dynamic Spectrum Access. IEEE Signal Proc. Mag. 79(3), 79–89 (2007)
2. Akyildiz, I.F., Melodia, T., Chowdhury, K.R.: A Survey on Wireless Multimedia Sensor Networks. Computer Networks 51(4), 921–960 (2007)
3. Pioro, M., Medhi, D.: Routing, Flow, and Capacity Design in Communication and Computer Networks. Morgan Kaufmann (2004)
4. Coersmeier, E., Hueske, K., Hoffmann, M., Leder, F., Martini, P., Bothe, H.: Combining Cognitive Radio and Software Radio Approach for Low Complexity Receiver Architecture. In: Proc. International Symposium on Advanced Radio Technologies (February 2007)
5. Lu, D., Huang, X., Zhang, W., Fan, J.: Interference-aware spectrum handover for cognitive radio networks. Wireless Communications and Mobile Computing (to appear)
6. Shokri-Ghadiklaei, H., Sheikholeslami, F., Nasiri-Kenari, M.: Distributed Multiuser Sequential Channel Sensing Schemes in Multichannel Cognitive Radio Networks. IEEE Trans. Wireless Commun. 12(5), 2055–2067 (2013)
7. Chen, Y., Chen, J.: A Fast Subcarrier, Bit, and Power Allocation Algorithm for Multiuser OFDMA-Based Systems. IEEE Trans. Vehicular Tech. 57(2), 873–881 (2008)
8. Le, L.B., Hossain, E.: Resource Allocation for Spectrum Underlay in Cognitive Radio Networks. IEEE Trans. Wireless Commun. 7(12), 5306–5315 (2008)
9. Eichfelder, G.: Adaptive Scalarization Methods in Multiobjective Optimization. Springer (2008)
10. Ehsan, N., Cruz, R.L.: On the Optimal SINR in Random Access Networks with Spatial Reuse. In: Proc. IEEE Conference on Information Sciences and Systems (March 2006)
11. Boyd, S., Kim, S.J., Vandenberghe, L., Hassibi, A.: A Tutorial on Geometric Programming. Optimization and Engineering 8(1), 67–127 (2007)

TTL-Based UDP Hole Punching Scheme in SIP Network

KwangHee Choi[1], Ki-Sik Kong[2], Kwang-Sik Chung[3],
Doo-Soon Park[4], and Joon-Min Gil[5,*]

[1] LG Uplus, Daejeon 305-350, Korea
theidea@lguplus.co.kr
[2] Dept. of Multimedia, Namseoul University, Chungnam 331-707, Korea
kskong@nsu.ac.kr
[3] Dept. of Computer Science, Korea National Open University, Seoul 110-791, Korea
kchung0825@knou.ac.kr
[4] Dept. of Computer Software Eng., Soonchunhyang University, Chungnam 336-745, Korea
parkds@sch.ac.kr
[5] School of IT Eng., Catholic University of Daegu, Gyeongbuk 712-702, Korea
jmgil@cu.ac.kr

Abstract. In this paper, we propose the low Time-To-Live (TTL)-based UDP hole punching scheme to reduce the load on the SIP registrar for Network Address Translator (NAT) traversal. For this purpose, the low TTL value is determined by executing 'traceroute' for the server reflexive IP address obtained through the SIP registration. In the proposed scheme, since the UDP hole punching request with low TTL value will never reach the registrar, the load on the registrar can be significantly alleviated. We analyze the proposed scheme using the Markov chain model. Numerical results indicate that the proposed scheme can significantly alleviate the load on the registrar, compared to the existing SIP scheme.

Keywords: NAT traversal, TTL, UDP hole punching, SIP.

1 Introduction

Session Initiation Protocol (SIP) is popular for establishing media sessions [1]. There are many devices and servers which are interoperable to SIP. Since a Network Address Translator (NAT) has been widely deployed in the home network, the SIP User Agents (UAs) like SIP phones are very often located behind NATs even if the Internet service provider gives a global IP address. In real network, LG Uplus, one of the major carriers in Korea, uses the access point (AP) working as a NAT for VoIP services [2]. It means that most Wi-Fi phones should frequently send REGISTER message to the registrar for NAT traversal (e.g., within 60 sec) [3]. Therefore, for such frequent registrations, the SIP registrar tends to suffer from the heavy load. To overcome this problem, we propose the low Time-To-Live (TTL)-based UDP hole punching

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_21, © Springer-Verlag Berlin Heidelberg 2014

scheme. In the proposed scheme, since the UDP hole punching request with low TTL value will never reach the registrar, the load on the registrar can be significantly alleviated. A recent work assumes that most NATs will not forward 'Internet Control Message Protocol (ICMP) TTL Exceeded' message back to an internal host [4]. But the assumption is not reasonable since the default setting of NAT device must allow 'ICMP TTL Exceeded' message to be originated [5]. In this paper, the assumption is not applicable. We determine the low TTL by executing 'traceroute' for the server reflexive IP address, currently an IP address owned by the NAT, obtained through the SIP registration procedure. To the best of our knowledge, this is the first work which supports the low TTL-based UDP hole punching in SIP environment.

2 Proposed Scheme

2.1 Reference Architecture

We first describe the reference architecture, as depicted in Fig.1. The SIP registrar has globally routable IP address and is connected to the main Internet, but most UAs are located behind one or multiple levels of NATs. The home NATs mean the AP or the home router which are essential components for triple-play-service (e.g., VoIP, IPTV and broadband services). In this paper, we focus on home NATs since major ISPs in Korea give global IP addresses to their subscribers. To explain our proposed scheme easily, we borrow the terms from ICE [6] as follows:

- Host IP address: an IP address on the host. If the host is located behind a NAT, the host IP address will be a private IP address.
- Server Reflexive IP address: an IP address which is allocated by the outermost NAT for an UA when it sends a packet through the NAT to a SIP registrar. The server reflexive IP address can be learned by SIP registrar.

In Fig. 1, the host IP address and the server reflexive IP address for the UA are A and B, respectively.

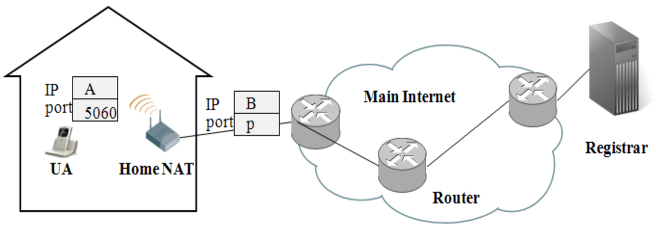


Fig. 1. Reference architecture

2.2 Low TTL-Based UDP Hole Punching Scheme

The key idea is that the UA determines its low TTL value by executing 'traceroute' for the server reflexive IP address, defined as an IP address which is allocated by the outermost NAT which is the farthest away from itself when it sends a packet through the NAT to the SIP registrar, obtained through its SIP registration procedure. The UA can obtain its server reflexive IP address from ‘received’ parameter in the topmost ‘via’ header field value of the response message of REGISTER [7]. In order to detect the NAT, we use two properties observed as follows:

- Home NATs will forward ‘ICMP TTL Exceeded’ message back to the internal host according to RFC 5508 [5].
- Most home NATs will respond ‘ICMP Echo reply’ message when they receive ‘ICMP Echo request’ message with their own IP address as a destination IP address. But, some NATs will never respond.

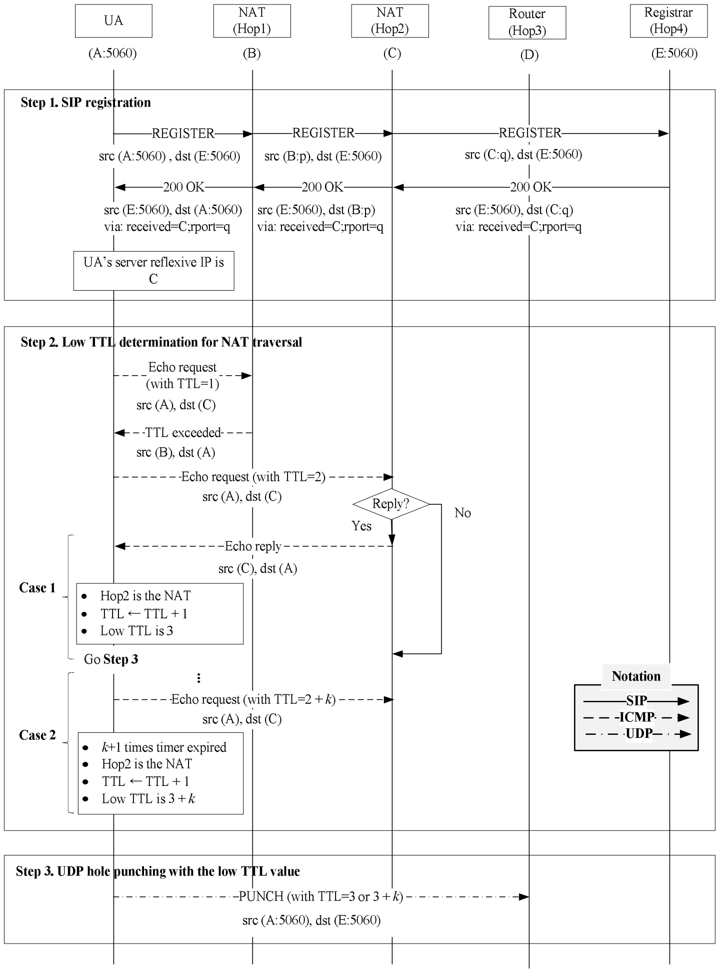


Fig. 2. The message flow of the proposed scheme

The proposed scheme is illustrated in Fig. 2. The UA's IP address is A, and the IP addresses for Hop1, Hop2, Hop3 and Hop4 are 'B', 'C', 'D' and 'E', respectively. Hop1 and Hop2 are home NATs, such as an AP or a home router. Hop3 is a router, not a NAT. Hop4 is the registrar which has a global IP address. The UA is located behind double levels of NATs. The UA can obtain its own server reflexive IP address 'C' through the SIP registration mentioned above (**step 1**). The UA sends 'ICMP Echo request' message to its own server reflexive IP address 'C' as a destination IP address with TTL value, starting at 1. There are two cases. The **case 1** is that the NAT responds 'ICMP Echo reply' message. The **case 2** is that the NAT does not respond 'ICMP Echo reply' message, and the timer is expired. In the **case 1**, the UA clearly can know that Hop2 is the outermost NAT which is the farthest away from itself. Therefore, the UA determines that the low TTL value for UDP hole punching is 3, that is, current TTL value + 1. However, some NATs will not send 'ICMP Echo reply' message in case that they are configured with an option to disable 'Echo server'. If the UA does not receive any 'ICMP Echo reply' and its timer expires for $k + 1$ times Echo requests (**case 2**), it considers Hop2 as a NAT, where k is equal to or greater than 0. The low TTL value is determined by $3 + k$, that is, current TTL value + 1 (**step 2**). The UA in the **case 1** or the **case 2** periodically sends a UDP packet named 'PUNCH' with TTL = 3 or $3 + k$, respectively (**step 3**).

3 Performance Analysis

The performance can be measured as the number of registration messages on the registrar. We evaluate the performance of the proposed scheme and the existing SIP scheme using the Markov chain model. To develop the Markov chain model for the performance evaluation, we make the following assumptions:

- The arrival process for UAs follows Poisson distribution with rate of λ.
- The sojourn time of UAs follows exponential distribution with the rate of mean of $1/\mu$.

To find the steady state distribution, a two-dimensional Markov chain is required. However, in this paper, we use an approximated method introduced in [8] for simplicity of analysis. The offered UA load to the registrar (ρ) is λ/μ.

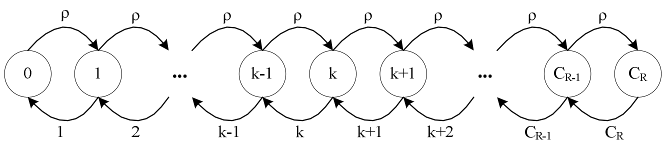


Fig. 3. Markov chain model

Fig. 3 shows the approximated one-dimensional Markov chain model. The capacity of the registrar (C_R) is represented by the maximum number of UAs that it can serve. The transition rates are as follows:

$$\begin{aligned} q(i,i+1) &= \rho \qquad && (0 \le i < C_R) \\ q(i+1,i) &= i+1 \qquad && (0 \le i < C_R). \end{aligned} \tag{1}$$

where $q(i,i+1)$ and $q(i+1,i)$ are the transition rate from state i to state $i+1$ and the transition rate from state $i+1$ to state i, respectively. In this Markov chain, state i refers to the number of UAs registered. The steady state probability (p_k) can be obtained as follows:

$$p_k = \frac{\rho^k / k!}{\sum_{n=0}^{C_R} \rho^n / n!} \qquad (0 \le k \le C_R). \tag{2}$$

In the existing SIP scheme, each MN registers with the registrar as soon as it starts up, and it will send the REGISTER message to the registrar at every UDP hole punching timer value (τ_H) of its NAT. However, in the proposed scheme with the low TTL-based scheme, the registrar will only receive REGISTER messages within a SIP registration expiration timer value (τ_R) since low TTL-based hole punching messages cannot be transmitted to the registrar. Usually, τ_R is much longer than τ_H since the default expiration for SIP registration is 3600 seconds ($\tau_R >> \tau_H$). Therefore, the numbers of registrations on the registrar of the existing SIP scheme and the proposed scheme, denoted as N_{SIP} and N_{PRP}, can be calculated as follows:

$$N_{SIP} = \sum_{k=0}^{C_R} p_k \cdot k \cdot \left(1 + \left[\frac{1}{\tau_H \cdot \mu}\right]\right), \qquad N_{PRP} = \sum_{k=0}^{C_R} p_k \cdot k \cdot \left(1 + \left[\frac{1}{\tau_R \cdot \mu}\right]\right). \tag{3}$$

where brackets mean Gauss' notation.

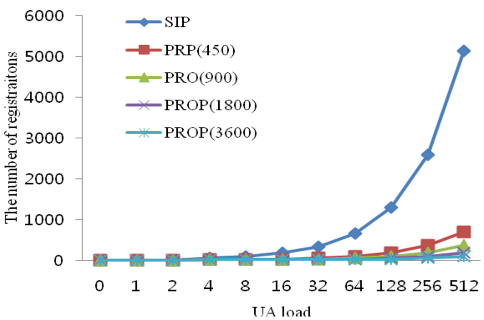


Fig. 4. The effect of UA load on the number of registrations

For numerical analysis, C_R is assumed to be 30. The arrival rate of UAs is set to 1/20 (1/sec). The UDP hole punching timer value of the NAT is set to 60 seconds, and the SIP registration expiration timer values are set to 450, 900, 1800, and 3600 seconds. The sojourn time is changed from 20 seconds to 10240 seconds, in order to vary the offered load. Fig. 4 shows the comparison between the existing SIP scheme and the proposed scheme in terms of the number of registrations. We investigate the effect of the number of the registrations according as the UA load is ranging from 2 to 512. As the UA load increases, the performance of the proposed scheme becomes more effective, because the proposed scheme can avoid the UDP hole punching load on the registrar. Recently, smart devices which can support SIP-based applications are widely popular. It means that the offered UA load to the registrar tends to increase. Considering such trends, therefore, the proposed scheme is expected to be more effective.

4 Conclusion

In this paper, we proposed a low TTL-based UDP hole punching scheme in SIP environment. The proposed scheme could reduce the SIP registration request load on the registrar remarkably since the UDP hole punching message is passed to the next hop of the NAT, not the registrar. In our future work, we will estimate real results of the proposed scheme in our SIP network.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A4A01015777).

References

1. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: Session Initiation Protocol. IETF RFC 3261 (2002)
2. Kim, H.: Business Models for the Free Public Wireless LAN Service. In: ICIN 2010, pp. 1–5 (2010)
3. Koski, P., Ylinen, J., Loula, P.: The SIP-Based System Used in Connection with a Firewall. In: AICT-ICIW 2006, pp. 203–208 (2006)
4. Biggadike, A., Ferullo, D., Wilson, G., Perrig, A.: NATBLASTER: Establishing TCP Connections Between Hosts Behinds NATs. In: ACM SIGCOMM Asia Workshop (2005)
5. Srisuresh, P., Ford, B., Sivakumar, S., Guha, S.: NAT Behavioral Requirements for ICMP. IETF RFC 5508 (2009)
6. Rosenberg, J.: Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols. IETF RFC 5245 (2010)
7. Rosenberg, J., Schulzrinne, H.: An Extension to the Session Initiation Protocol (SIP) for Symmetric Response Routing. IETF RFC 3581 (2003)
8. Fang, Y., Zhang, Y.: Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks. IEEE Transactions on Vehicular Technology 51(2), 371–382 (2002)

An Empirical Study on the Quality Assessment of the VoIP Service over Wireless Mobile Networks

Kwang-Yeol Yoon and Beomjoon Kim*

Department of Electronic Engineering, Keimyung University, Daegu, 704-701, Korea
{yoonky,bkim}@kmu.ac.kr

Abstract. This paper discusses the service quality of packet-based voice service provided over wireless mobile networks such as wireless broadband (WiBro) and high speed downlink packet access (HSDPA) systems. Using measurement software, a large scale of experiment has been conducted to measure the actual quality of the voice service over both wireless mobile networks. Based on the results from the experiment, the quality of the voice service is supposed to be quite good. Through further experiment, the quality degradation over a radio channel leads to the increase in delay and the subsequent quality degradation of the voice service.

Keywords: mobile VoIP service, voice quality measurement software, E-Model, mean opinion score (MOS).

1 Introduction

The packet-based voice service such as voice over IP (VoIP) service over a wired network [1] has become popular now because of its low cost. In providing the voice service over a wireless mobile network, however, one of the challenges that should be addressed is how to support and guarantee its service quality [2], [3]. Not only the bandwidth provided by a wireless channel is small, several characteristics should be considered; e.g., the time-varying channel qualities over a wireless channel, mobility-related processes like handoff, and etc.

For the past few years, we have studied to find the answers for the following two questions:

- What is the current service quality of voice service over wireless mobile networks?
- What level of radio channel quality provokes the degradation of the service quality of the voice service over wireless mobile networks?

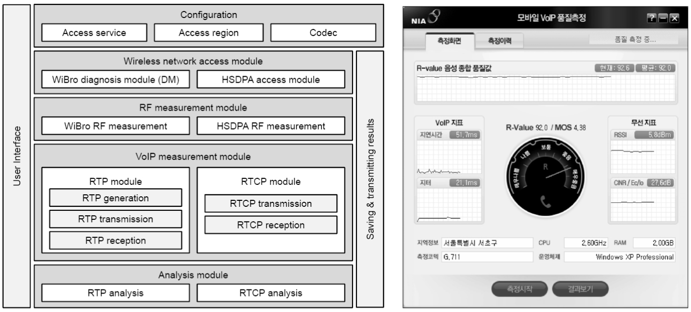
And the answers are presented in the remainder of this paper.

* Correspoding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_22, © Springer-Verlag Berlin Heidelberg 2014

2 The Quality Measurement Software

In order to measure the voice quality in this study, we use the software developed in our previous research [4]. Fig. 1-(a) shows the logical architecture of the software running on the user equipment's side.



(a) the logical architecture (b) a screen shot of the output

Fig. 1. The voice quality measurement software

The software consists of the measurement function, the reporting function, and the user interface. It supports the access to two wireless mobile networks that are available now; one is wireless broadband (WiBro) that is based on IEEE 802.16 standard and well-known as WiMAX and the other one is high speed downlink packet access (HSDPA) that is a 3G mobile system.

Table 1. The selected service quality metrics

	Quality metrics	
	WiBro	HSDPA
RF quality metrics	RSSI (received signal strength indicator)	
	CINR (carrier to interference noise ratio)	Ec/Io (energy per chip over the interference noise)
Network quality metrics	Bandwidth, one-way delay, jitter, packet loss rate	
VoIP quality metrics	R-Score, MOS (mean opinion score)	

Considering the communication architecture in which voice packets are delivered, the measurement function has been implemented by the RF measurement module and the VoIP measurement module. A few quality metrics are selected in consideration of their importance and summarized in Table. 1. The RF quality metrics are measured by the RF measurement module while the network and VoIP quality metrics by the VoIP measurement module.

The wireless quality metrics are measured based on the information obtained through the standard interface and commands provided by the modem manufacturer. The RFCs are referred to for the algorithm to measure the network quality metrics;

i.e., one-way delay [6], packet loss ratio [7], and jitter [8]. For the VoIP quality metrics, we follow the E-Model [5] that derives R-Score and mean opinion score (MOS) using the measured values of network quality metrics. Fig. 1-(b) shows a screen shot generated by the software with the measured values.

3 Experiments

3.1 Experiment 1

In order to investigate the current status, we have conducted an experiment to measure the VoIP service quality over wireless mobile networks. The measurement software runs on the user equipment that is implemented by a laptop. The voice traffic for test generated by a server is delivered to the user equipment through actual WiBro and HSDPA mobile networks. Two codecs are considered; one is G.711 that requires 64kbps and the other one is G.729 that requires 8kbps with compression.

Reflecting various transmission characteristics over a wireless link, the measurement is performed under the following five conditions: indoor-stationary (E1), indoor-walk (E2), outside-stationary (E3), outside-walk (E4), and outside-moving (E5). A measurement period lasts for 300 seconds. A data sample is made from averaging 60 values measured in every 5 seconds during the measurement period. For each scenario, 200 samples are obtained and the results are shown in Fig. 2.

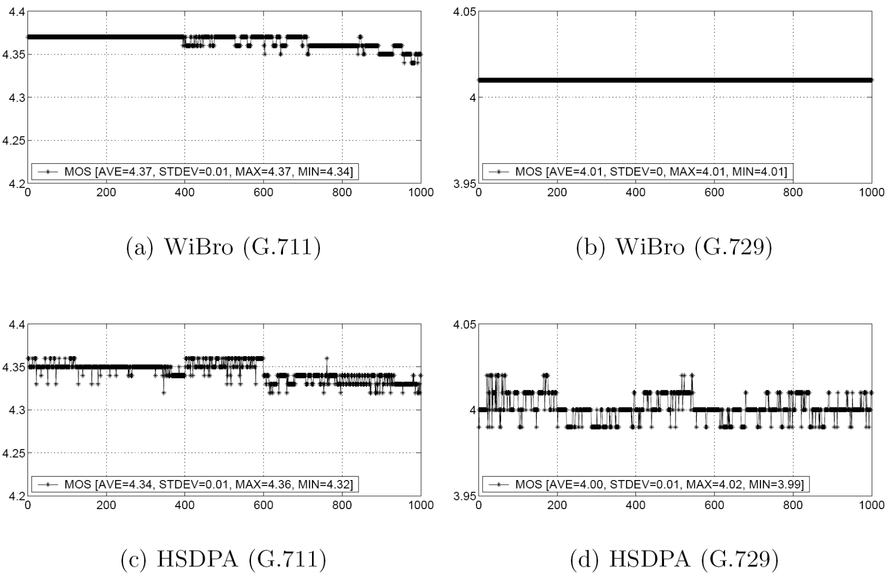


Fig. 2. The measured MOS (x-axis: number of sample, y-axis: the measured value)

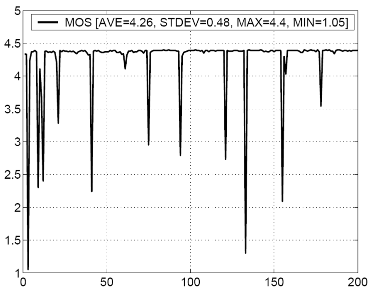
During the measurement, the bandwidth measured at the user equipment ranges from 5.93 to 7.82 for WiBro and from 2.09 to 3.18 for HSDPA in average. Considering that even G.711 that adopts no compression requires only 64 kbps, the provided bandwidth is quite sufficient to accommodate the voice traffic. Although the MOS values in four diagrams decrease a little for the samples 800 through 1000 comprising E5, the overall value shows better than 4.0 that corresponds to 'toll-quality'. Note that the highest value of MOS with G.729 is no better than G.711 because of compression.

3.2 Experiment 2

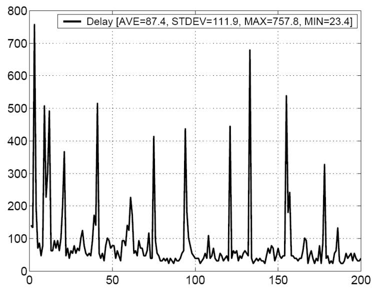
In Experiment 2, we have conducted a sort of "stress-test" to figure out the RF channel quality that provokes the voice quality degradation. For this experiment, it should be noted that a radio channel emulator is used to adjust the radio channel quality artificially. We have also determined three configurations applied to the radio channel emulator as in Table 2. For each configuration, 200 samples are obtained for WiBro and HSDPA in terms of G.711 and G.729 in respect.

Table 2. The RF channel configuration for experiment 2

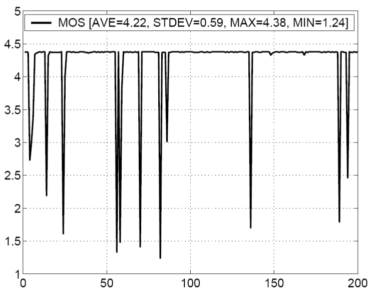
	WiBro	HSDPA
Configuration 1 (C1)	RSSI -70(dBm), CINR 16(dB)	RSSI -65(dBm), Ec/Io -5 (dB)
Configuration 2 (C2)	RSSI -75(dBm), CINR 14(dB)	RSSI -65(dBm), Ec/Io -10(dB)
Configuration 3 (C3)	RSSI -80(dBm), CINR 12(dB)	RSSI -65(dBm), Ec/Io -15(dB)



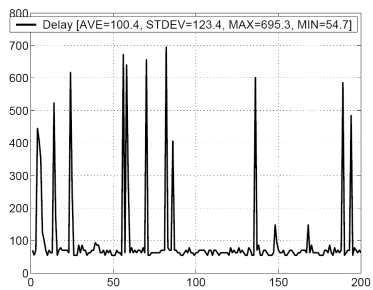
(a) WiBro, MOS, C3, G.711



(b) WiBro, Delay(msec), C3, G.711



(c) HSDPA, MOS, C3, G.711



(d) HSDPA, Delay(msec), C3, G.711

Fig. 3. The measured MOS and delay for WiBro and HSDPA
(x-axis: number of sample, y-axis: the measured value)

Fig. 3 shows the measured results for WiBro and HSDPA. Since every measured value of MOS maintain above 4 for C1 and C2, we have included the results only for C3. In Fig. 3-(a), it is observed that sometimes MOS decreases very sharply and deeply. The reason for such decrease can be found in the measured delay values shown in Fig. 3-(b). It can be shown that the measured delay value increases very high at the same number of sample where MOS decreases. Almost similar results are observed for HSDPA as shown in Fig. 3-(c) and 3-(d). For G.729, the MOS values in every measurement always stay above 4, which can be explained that it requires only 8 kbps.

4 Conclusion

In this paper, we have investigated the service quality of VoIP service over wireless mobile networks. Based on the measurement results from our experiments, it can be concluded that WiBro and HSDPA networks provide sufficient bandwidth so that the voice quality is quite good and the degradation over a radio channel leads to the increase in delay and the subsequent degradation of the voice quality.

References

1. Markopoulou, A.P., Tobagi, F.A., Karam, M.J.: Assessing the Quality of Voice Communications over Internet Backbones. IEEE/ACM Transactions on Networking 11(5) (2003)
2. Scalabrino, N., Pellegrini, F.D., Riggio, R., Maestrini, A., Costa, C., Chlamtac, I.: Measuring the Quality of VoIP Traffic on a WiMAX Testbed. In: Proceeding of Tridentcom (2007)
3. Han, M., Lee, Y.-S., Moon, S., Jang, K., Lee, D.: Evaluation of VoIP Quality over WiBro. In: Claypool, M., Uhlig, S. (eds.) PAM 2008. LNCS, vol. 4979, pp. 51–60. Springer, Heidelberg (2008)
4. Kim, C.-C., Kim, B.: A Software Implementation for Quality Management of Mobile VoIP Services. In: Kim, T.-H., Kang, J.-J., Grosky, W.I., Arslan, T., Pissinou, N. (eds.) MulGraB/BSBT/IUrC 2012. CCIS, vol. 353, pp. 73–80. Springer, Heidelberg (2012)
5. ITU, The E-Model, A Computational Model for Use in Transmission Planning: ITU-T Recommendation G.107 (1998)
6. Almes, G., Kalidindi, S., Zekauskas, M.: A One-way Delay Metric for IPPM. RFC 2679 (1999)
7. Almes, G., Kalidindi, S., Zekauskas, M.: A One-way Packet Loss Metric for IPPM. RFC 2680 (1999)
8. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (2003)
9. ITU-T, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). ITU-T Recommendation G.729 (2007)

Classification of Painting Genres Based on Feature Selection

Yin-Fu Huang and Chang-Tai Wang

Department of Computer Science and Information Engineering
National Yunlin University of Science and Technology,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
{huangyf,m10017032}@yuntech.edu.tw

Abstract. In this paper, a painting genre classification system is proposed. Four feature descriptors about the color and texture defined in the MPEG-7 specification, which are more against painting characteristics, are extracted from data sets. Then, we use a self-adaptive harmony search algorithm to select relevant features (or a local feature set) to train each one-against-one SVM classifier. Finally, a majority voting strategy on N(N-1)/2 prediction results would determine their respective genres of paintings. The experimental results show that the overall accuracy reaches 69.8%, and this demonstrates more precise features can be selected for each pair of genres to get better classification results.

Keywords: Painting genre classification, MPEG-7 specification, harmony search algorithm, majority voting strategy.

1 Introduction

Currently, most image classification systems rely on visual features to achieve satisfactory results. In general, the visual features are divided into color features, texture features, and even shape features extracted from the characteristics in the frequency domain. Besides, many bag-of-features [8] have been proposed for uses to build their own feature datasets. Here, we also use the bag-of-feature based on MPEG7 specification; i.e, four feature descriptions of color and texture are adopted to generate the initial feature set for painting genre classification.

In order to gather up the enough knowledge for classification, a generated feature set commonly contains the abundant information with some probably redundant features. For solving this problem, dimensionality reduction techniques are frequently employed and they can be classified into two approaches. One kind is to transform the matrix of a feature set from a high-dimensional space to a lower-dimensional space through the linear combination of matrix using the techniques such as principal component analysis (PCA) [6], non-negative multi-linear principal component analysis (NMPCA), et al. Another kind is called the feature selection which finds an optimum subset from the original feature set using the search algorithms such as genetic algorithm (GA) [7], ant colony optimum (ACO) [3], harmony search (HS) [4], et al. Both these two approaches can effectively reduce the dimensions of a feature set.

For the prediction, neural networks, support vector machine (SVM) [1, 2], multilayer perceptron (MLP) [9], and linear regression [10] are often used as classifiers.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_23, © Springer-Verlag Berlin Heidelberg 2014

Here, an SVM classifier is adopted since, in general, it presents better performances than other classifiers [2] while a kernel function and parameters are appropriately chosen.

The remainder of this paper is organized as follows. In Section 2, we present the system architecture. In Section 3, the SAHS algorithm and correlation measuring method are proposed to derive an optimum feature subset from the original feature set. In Section 4, we present the experimental results on different selection strategies. Finally, we make conclusions in Section 5.

2 System Overviews

In this paper, we propose a painting genre classification system consisting of a training phase and a test phase. As shown in Fig. 1, in the training phase, four feature descriptors about the color and texture defined in the MPEG7 specification are extracted from a training data set first; i.e., ColorLayout descriptor, ColorStructure descriptor, EdgeHistogram descriptor, and HomogeneousTexture descriptor including a total of 186 dimensions. Then, we use a meta-heuristic optimization algorithm called Self-Adaptive Harmony Search (i.e., SAHS) algorithm [5] to select relevant features (or a local feature set) to train each one-against-one SVM classifier. For N painting genres to be classified, we would have N(N-1)/2 SVM classifiers to be trained; in other words, N(N-1)/2 local feature sets should be generated to train N(N-1)/2 SVM classifiers.

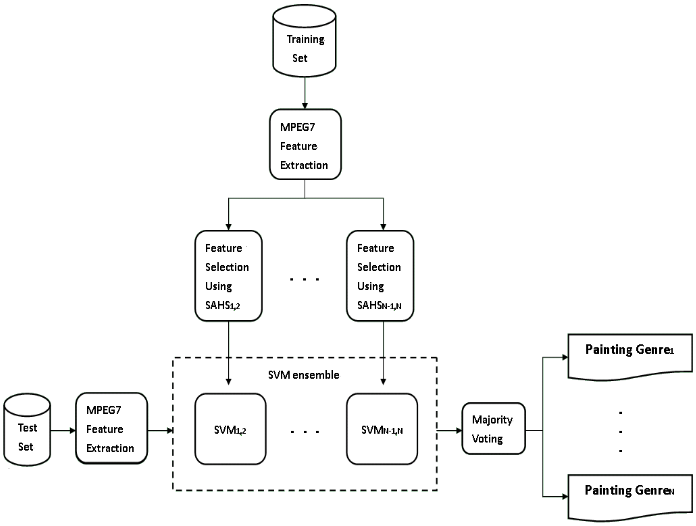


Fig. 1. System architecture

In the test phase, we also extract the same features, as mentioned, from a test data set. Then, each local feature set generated using the SAHS algorithm is fed into the corresponding one-against-one classifier in a SVM ensemble model. Finally, a majority voting strategy on N(N-1)/2 prediction results would determine their respective genres of test data. The feature selection model used here enables ambiguous genres to be classified more precisely.

3 Feature Selection

In our work, the feature selection model consists of two parts: the self-adaptive harmony search (i.e., SAHS) algorithm and the objective function [11], as shown in Fig. 2. Once the original feature set is given, the SAHS algorithm begins to iteratively search a better solution that would be evaluated later by the objective function. Finally, the best solution will be output as the final feature subset.

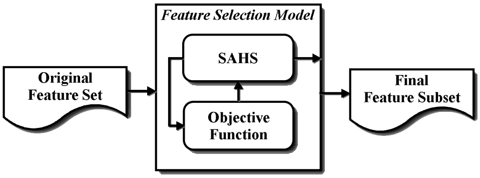


Fig. 2. Feature selection model

When each new harmony symbolizing a selected feature subset is generated from the SAHS algorithm, the relative correlation (or the objective function) is used to evaluate the performance of the selected feature subset. The correlations of the selected feature subset are conducted in two phases; one is the intra-correlation evaluating the mutual correlation between features within the subset, and another is the inter-correlation comparing each feature inside the subset with the corresponding class. If a subset has better performance, it must possess the property of lower intra-correlation and higher inter-correlation. The lower intra-correlation means the features within the subset are relevant, whereas the higher inter-correlation means each feature within the subset is discriminative for the corresponding class.

4 Experimental Results

In this study, our dataset includes 6 painting genres; i.e., cubism (72 pieces), fauvism (74), impressionism (74), naïve art (49), pointillism (71), and realism (71). In the experiments, about two-thirds of the dataset are used for training, and one-third of the dataset is used for test. We adopt the well-known LIBSVM developed by Chang and Lin as an SVM classifier. The kernel function used here is RBF (i.e., Radial Basis Function) since it is more accurate and effective than the other kernel ones. Then, we find the optimum parameter values are $\gamma = 2^{-6}$ and $C = 2^5$. Moreover, the feature vector is normalized in the range [-1, 1].

4.1 Original Feature Set

In this experiment, the original feature set totally including 186 features is used as the feature vector. The classification result is illustrated in Table 1 and the overall classification accuracy is around 64.3%. This result also demonstrates that the original feature set is unsuitable for some genre like fauvism since the recall is less than 50%. It means that the disturbance of some unsuitable features in the original feature set results in disabling to clearly discriminate these genres.

Table 1. Confusion matrix on the original feature set

	Cub	Fau	Imp	Naï	Poi	Rea	*Recall*
Cubism	16	3	1	0	0	2	72.7%
Fauvism	9	9	1	1	1	1	40.9%
Impressionism	6	2	13	0	1	0	59.1%
Naïve art	1	1	0	13	1	0	81.3%
Pointillism	0	2	4	0	15	1	68.2%
Realism	3	0	4	0	0	15	68.2%
Precision	45.7%	52.9%	56.5%	92.9%	83.3%	78.9%	64.3%

4.2 Global Selection Strategy with SAHS

In this experiment, the global feature set for 6 genres is selected, which consists of 54 features. For the classification result illustrated in Table 2, the overall accuracy is around 67.5%, and increases by 3.2% when compared with the result in Table 1. Almost all precision and recall are improved, especially for the genre like fauvism. This result demonstrates that the feature selection model is effective to remove the unsuitable features for classification and only 29% of the original features are used to get better performances.

Table 2. Confusion matrix by the global selection strategy

	Cub	Fau	Imp	Naï	Poi	Rea	*Recall*
Cubism	16	3	2	0	0	1	72.7%
Fauvism	6	12	2	0	1	1	54.5%
Impressionism	4	1	15	1	1	0	68.2%
Naïve art	2	1	0	12	1	0	75.0%
Pointillism	1	1	3	0	16	1	72.7%
Realism	3	0	5	0	0	14	63.6%
Precision	50.0%	66.7%	55.6%	92.3%	84.2%	82.4%	67.5%

4.3 Local Selection Strategy with SAHS

In this experiment, for 6 painting genres, 15 one-against-one local feature sets are generated for classification and the number of features in each set is illustrated as Table 3. This table is symmetric and the numbers of features are much less than those (i.e., 54 features) in the global feature set. For the classification result illustrated in Table 4, the overall accuracy is around 69.8%, and increases by 2.3% when compared with the result in Table 2. This result demonstrates that, without considering the other genres, more precise features can be selected for each pair of genres in order to get better classification results than the global feature set.

Table 3. Number of features in each local feature set

Cubism	Cub	Fau	Imp	Naï	Poi	Rea
Fauvism	*	39	38	31	30	36
Impressionism		*	32	34	28	27
Naïve art			*	38	32	32
Pointillism				*	29	28
Realism					*	33
Cubism						*

Table 4. Confusion matrix by the local selection strategy

	Cub	Fau	Imp	Naï	Poi	Rea	*Recall*
Cubism	17	2	2	0	0	1	77.3%
Fauvism	7	12	0	2	0	1	54.5%
Impressionism	4	1	16	0	1	0	72.7%
Naïve art	2	1	0	12	1	0	75.0%
Pointillism	0	2	3	0	16	1	72.7%
Realism	2	0	5	0	0	15	68.2%
Precision	53.1%	66.7%	61.5%	85.7%	88.9%	83.3%	69.8%

Here, the qualitative comparisons between our method and the other existing ones are illustrated in Table 5. Although the datasets used could be different from each other, they always use the Google search engine to collect their datasets. Only two methods including ours use dimensionality reduction techniques to filter out irrelevant features. Then, the various classifiers are used for painting genre classification. Here, we use the feature selection model based on the local selection strategy to help us achieve the second best performances among all methods.

Table 5. Comparisons among all methods

	J. Zujovic et al. [12]	**M. Culjak et al. [2]**	**J. Zujovic et al. [12]**	**M. Culjak et al. [2]**	**Ours**
Dataset	Google search & CARLI collections	Google search & Artlex database	Google search & CARLI collections	Google search & Artlex database	Google search & Artlex database
Original feature	-	68	-	68	186
Genres	5	6	5	6	6
Dim. reduction	-	-	-	-	SAHS
Classifier	SVM	ANN	AdaBoost	SMO	SVM
Accuracy	57.8%	56.6%	68.3%	60.2%	69.8%

5 Conclusions

In this paper, we propose a painting genre classification system. The original feature set includes four feature descriptors about the color and texture defined in the MPEG7 specification. By employing the SAHS (self-adaptive harmony search) algorithm on the original feature set, the feature selection model can effectively find the optimum feature subsets for corresponding painting genres. For the experimental results using the SVM classifier, the local selection strategy can present better performances than the global selection strategy. In other words, the local selection models can derive more relevant features than the global selection model. In summary, the experimental results demonstrate that our method is more effective than the other existing ones.

Acknowledgments. This work was supported by National Science Council of R.O.C. under grant NSC100-2218-E-224-011-MY3.

References

1. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001), `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
2. Culjak, M., Mikuš, B., Jež, K., Hadjic, S.: Classification of art paintings by genre. In: Pro. the 34th International Convention on Information and Communication Technology, Electronics and Microelectronics, pp. 1634–1639 (2011)
3. Deriche, M.: Feature selection using ant colony optimization. In: Proc. the 6th International Multi-Conference on Systems, Signals and Devices, pp. 1–4 (2009)
4. Diao, R., Shen, Q.: Two new approaches to feature selection with harmony search. In: Proc. the IEEE International Conference on Fuzzy Systems, pp. 1–7 (2010)
5. Huang, Y.-F., Wang, C.-M.: Self-adaptive harmony search algorithm for optimization. Expert Systems with Applications 37(4), 2826–2837 (2010)
6. Jin, X., Bie, R.: Random forest and PCA for self-organizing maps based automatic music genre discrimination. In: Proc. the International Conference on Data Mining, pp. 414–417 (2006)
7. Lanzi, P.L.: Fast feature selection with genetic algorithms: a filter approach. In: Proc. the IEEE International Conference on Evolutionary Computation, pp. 537–540 (1997)
8. Martinez, J.M., Koenen, R., Pereira, F.: MPEG-7: the generic multimedia content description standard, part 1. IEEE Multimedia 9(2), 78–87 (2002)
9. Mitra, V., Wang, C.J.: A neural network based audio content classification. In: Proc. the International Joint Conference on Neural Networks, pp. 1494–1499 (2007)
10. Mokadem, A., Charbit, M., Chollet, G., Bailly, K.: Age regression based on local image features. In: Proc. the 4th Pacific-Rim Symposium on Image and Video Technology, pp. 88–93 (2010)
11. Silviu, G.: Information Theory with Applications. McGraw-Hill (1977)
12. Zujovic, J., Gandy, L., Friedman, S., Pardo, B., Pappas, T.N.: Classifying paintings by artistic genre: an analysis of features & classifiers. In: Proc. the IEEE International Workshop on Multimedia Signal Processing, pp. 1–5 (2009)

Mining Domain-Dependent Noun Opinion Words for Sentiment Analysis

Zhe Qi and Mingchu Li

School of Software, Dalian University of Technology
Dalian 116620, Liaoning, P.R. China
meloqizhe@gmail.com, mingchul@dlut.edu.cn

Abstract. Mining domain-dependent opinion words is an important problem in opinion mining, which facilitates the extraction of opinions from text. However, existing work pay close attention to adjectives and verbs, only limited work focuses on noun and noun phrases. In our work, we use opinion mining method based on linguistic analysis to identify and extract objective noun product features that imply positive or negative opinions in specific domains. Firstly, we presented a method for automatic extraction of noun features. The extraction method used linguistic frame to extract adverb-adjective-noun phrases. Subsequently, for adverb-adjective-noun phrases, we proposed a general additive model to compute sentiment scores for noun features. Then, by implementing a statistic method and pruning non-opinionated noun features, we determined objective noun features that imply sentiment in specific domains. At last, we conducted several experiments on English datasets to manifest the advantages of our scheme. Based on real-life datasets, the proposed method shows merits in accuracy and robustness. Moreover, it has promising recall and precision.

Keywords: noun features, opinion mining, specific domains, linguistic frame, general additive model, statistic method.

1 Introduction

On review websites and social networks, a growing corpus of online informal reviews is generated by non-experts every day. Millions of people express their opinions on a wide range of products, events and services by using opinion words. Opinion words are words that describe the emotion and attitude to certain opinion targets. Adjectives, verbs and nouns are main parts of opinion words, which are subjective or objective. The key difficulty in identifying such words is that opinions expressed by many of these words are objective and context or domain dependent. Tackling the difficulty is very challenging and critical for opinion mining in specific domains.

In general, opinion words are divided into two kinds: "pure" and "conditional" opinion words. The "pure" opinion words ("excellent", "boring", etc.) indicate subjective information and don't depend on the context or domain. Their polarities may not change to the opposite with the changing of domains. The "conditional" opinion words ("big", "size", etc.) indicate objective information and are domain-dependent.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_24, © Springer-Verlag Berlin Heidelberg 2014

When the domain or context is changed, in the case of “conditional” opinion words the polarity may change to the opposite one. Most of the work [1, 2, 3, 4, 5, 6, 8] for mining opinion words so far has employed “pure” opinion words and adjectives and verbs from “conditional” opinion words. However, nouns and noun phrases from “conditional” opinion words are not usually taken into consideration. For example, “Within half a month, a very big hole formed in the middle of the mattress”. Here, “hole” is a noun product feature and objective word. It describes the mattress’s quality and expresses a negative polarity. The opinion expressed by noun “hole” may not be found by most of the current techniques. Their involved sentences are also objective sentences and imply positive or negative opinions. These nouns, if not well extracted and utilized, can reduce the recall and precision of opinion word mining.

Existing approaches to opinion words extraction can be divided into two categories: corpus-based and dictionary-based methods. As dictionaries don’t contain domain-dependent information, dictionary-based methods [9-13] are not suitable for mining domain specific opinion words. The corpus-based paradigm [6-8] makes it possible to extract domain specific opinion word. However, corpus-based methods are mainly focused on adjectives and verbs. Riloff et al. [15] proposed a method to extract subjective nouns, but our attention is completely different to the existing methods because many nouns expressing opinion are objective nouns. Zhang and Liu [16] considered the task of extracting opinion words-nouns in specific domains firstly by using opinion lexicon compiled by Ding et al. [3]. Its shortcoming is the precision of extracted opinionated noun feature. Objective words or sentences that imply opinions are very difficult to recognize because their recognition typically requires the world knowledge and commonsense of the specific application domains. The common drawbacks of the aforementioned works can be summarized as follow: (1) identifying the objective noun words or sentences that imply opinions (2) improving the precision of extracted opinionated noun words. Suppose no action is going to be taken, the diverse problems for mining domain-dependent noun opinion words will occur.

Motivated by the above existing problems, in this paper we propose a novel method to identify and extract objective nouns product features which express opinions. In our work, first we extract adverb-adjective-noun phrases based on linguistic frame from product reviews. Then we obtain the noun product features from the extracted phrases. Subsequently, we compute sentiment scores for noun product features with a general additive model. Finally, by using a statistic approach, we also propose a robust solution for extracting noun product features that imply opinions.

The remainder of this paper is organized as follows: Section 2 presents our proposed method from three aspects: linguistic analysis, sentiment computation and choosing noun features that imply sentiment. Section 3 discusses a real-world scenario, applying our method to the real-life datasets and making the full evaluation. Finally, Section 4 draws conclusions about the work presented in this paper.

2 The Proposed Method

We use opinion mining approach based on linguistic frame and extract noun product features that imply opinions in specific application domains. To make this method concrete, we represent our method from three aspects: linguistic analysis, sentiment computation and choosing noun features that imply sentiment.

2.1 Linguistic Analysis

Our linguistic analysis uses the linguistic frame [17] instead of the flat structure of a surface string. The linguistic frame preserves linguistic structure by providing a hierarchical representation and encoding different layers of semantic dependencies.

```
{c cstatement
    :topic {q hole
              :quantifier "def"
              :modifier "fromed"
              :pred {p in :topic {q mattress
                                   : quantifier "def"}}}
    :adv "really"
    :pred {p adj_complement
             :pred {p adjective
                     :adv "very"
                     :pred {p quality :topic "big"}}}}
```

Fig. 1. Linguistic frame for "a hole formed in the middle of the mattress is really very big"

To make the concept of linguistic frame clear, we give an example: "Within half a month, a hole formed in the middle of the mattress is really very big". The linguistic frame of the sentence is shown in Fig. 1. We extract the head of topic sub-frame and the head of the predicate sub-frame that appear in the root layer of linguistic frame. Then, we find the adverbs "really" and "very" which modify head words. From the linguistic frame, we can easily extract adverb-adjective-noun phrases such as "really very big hole" by the constructed context free grammar rules.

2.2 Sentiment Computation

After extracting adverb-adjective-noun phrases, we intend to assign sentiment scores for noun product features. To calculate the scores, we devise a general additive model to calculate the degree of sentiment for noun features.

For each noun feature, we compute its score given by:

$$Score(n) = Score(adj) + Pol(adj) * Score(adv) + Pol(adj) * Score(neg) \tag{1}$$

where *Score(n)* represents the sentiment score for noun *n*, *Score(adj)* represents the sentiment score for adjective *adj*, Score(adv) represents the sentiment score for adverb *adv* and *Score(neg)* represents the sentiment score for negation *neg*. *Pol(adj)* represents the polarity of adjective *adj*.

As for each adjective, its score is given by:

$$Score(adj) = \frac{\sum_{i \in S} \frac{N}{n_{ri}} * r_i}{\sum_{ri} \frac{N}{n_{ri}}} \tag{2}$$

where S represents the set of appearances of adjective *adj*, r_i represents the user rating with each appearance of *adj*, *N* represents the number of reviews about the specific product, n_{ri} represents the number of reviews with rating r_i. As for adverbs and negations, we can get the sentiment score by using a slightly modified version of (2).

2.3 Choosing Noun Features That Imply Sentiment

By using sentiment computation, we can obtain the sentiment score for noun product features. To choose opinionated noun features, firstly we design two counters. One counter is used to count the number of appearance of the particular noun feature whose sentiment score is greater than 3 in each corresponding sentence. The other is used to count the number of appearance of the particular noun feature whose sentiment score is less than 3 in each corresponding sentence. Then, we compute the percentage of positive (or negative) opinions that the particular noun feature implies in all corresponding sentences.

Through experiments, we make an empirical assumption that if either the positive-opinionated opinion percentage or the negative-opinionated opinion percentage is significantly greater than 70%, we regard this noun feature as a noun feature implying an opinion [16]. A statistic approach is used to test for population proportion to conduct the significant test. We compute the Z-score statistic with one-tailed test.

$$Z = \frac{p - p_0}{\sqrt{\frac{p_0(1 - \mathrm{p}_0)}{n}}} \tag{3}$$

where p0 is a hypothesized value (0.7 in our experiment), p is the percentage of positive (or negative) opinions that the particular noun feature implies in all corresponding sentences, and n is the total number of all the review sentences associated with the particular noun feature. The statistical confidence level is set to 0.95. At this time, the corresponding Z value is -1.64. If Z score for a noun feature is greater than -1.64, the noun feature is an opinionated noun feature. Otherwise, the noun feature doesn't express any opinion.

3 Experiments

In this section we present a systematic evaluation of the proposed method on four different real-life datasets of product reviews. The domains corresponding to these four different datasets are Mattress, Drug, Router and Radio. Each dataset has 2000 product review sentences. The real-life datasets of product reviews are the similar to the datasets [16]. We obtained the datasets by JAVA crawler program.

To evaluate the performance of the proposed method, we compare it with a baseline method and a lexicon-based opinion mining method (LB) [16]. The baseline method determines the polarity of a noun feature only by its corresponding adjacent adjective. If the adjective is positive, the noun feature is positive. If the adjective is negative, the noun feature is negative. We can identify the noun features that imply opinions. LB method uses the lexicon-based opinion mining model to mine opinions in product reviews and finds the domain noun feature that express opinions.

Table 1. Results for Noun Features

	Baseline		LB		Proposed Method	
	Recall	Precision	Recall	Precision	Recall	Precision
Mattress	0.47	0.36	0.84	0.50	0.82	0.55
Drug	0.52	0.40	0.90	0.64	0.88	0.65
Router	0.70	0.32	0.76	0.42	0.75	0.48
Radio	0.80	0.28	0.85	0.48	0.80	0.60

Table 1 shows the results for extracted noun features by the linguistic frame. We use recall and precision to measure the performance. From Table 1, we note that the proposed method is the best on the precision. Compared with baseline, the proposed method gives that many opinionated noun features are not directly modified by subjective adjective opinion words but modified by objective adjectives. By comparison with LB, the precision of the proposed method increases 16% on average.

Table 2. Noun Features that Imply Positive Opinions

	Baseline	LB	Proposed Method
Mattress	0.30	0.45	0.50
Drug	0.35	0.60	0.60
Router	0.28	0.45	0.52
Radio	0.33	0.48	0.50

Table 3. Noun Features that Imply Negative Opinions

	Baseline	LB	Proposed Method
Mattress	0.40	0.48	0.55
Drug	0.32	0.64	0.64
Router	0.35	0.40	0.45
Radio	0	0	0

Table 2 shows the precision for noun features that imply positive opinions. Table 3 shows the precision for noun features that imply negative opinions. From Table 2 and Table 3, we note that our proposed method has the best precision for positive-opinionated and negative-opinionated noun features. The precision of the proposed method is 40% higher than the baseline method on average. Compared with the LB method, the precision of the proposed method is 13% higher on average. Note that there is no negative-opinionated noun features in Radio, because some specific domains may not contain positive-opinionated or negative-opinionated noun features.

4 Conclusions

This paper aims to extracting opinionated noun features in specific domains. Many noun features are objective words but express positive or negative opinions in some application domains. In this paper, a novel approach was presented to extract such opinionated noun features in product reviews. First, we presented a method for extracting noun product features automatically. The extraction method used linguistic frame to extract adverb-adjective-noun phrases. Subsequently, for adverb-adjective-noun phrases, a general additive model was proposed to compute the sentiment scores for noun features. Then, by implementing a statistic method and pruning non-opinionated noun features, we determined objective noun features that imply opinions in some application domains. Finally, our experimental results manifested the advantages of our scheme and showed that the proposed method is accurate and robust.

Acknowledgments. This work was supported by the National Science Foundation of Foundation of China under grant No. 61272173, 61100194 and the Fundamental Research Funds for the Central Universities under grants DUT12RC(3)80, DUT12ZD104 and DUT13LK38.

References

1. Breck, E., Choi, Y., Cardie, C.: Identifying Expressions of Opinion in Context. In: Proceedings of IJCAI 2007 (2007)
2. Kobayashi, N., Inui, K., Matsumoto, Y.: Extracting aspect-evaluation and aspect-of relations in opinion mining. In: Proceedings of EMLP 2007 (2007)
3. Ding, X., Liu, B., Yu, P.S.: A Holistic Lexicon-Based Approach to Opinion Mining. In: Proceedings of WSDM 2008, pp. 231–240 (2008)
4. Lu, Y., Zhai, C., Sundaresan, N.: Rated aspect summarization of short comments. In: Proceedings of WWW 2009, pp. 131–140 (2009)
5. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, 2nd edn. (2010)
6. Hatzivassiloglou, V., Kathleen, M.: Predicting the Semantic Orientation of Adjectives. In: Proceedings of ACL 1997, pp. 174–181 (1997)
7. Kanayama, H., Nasukawa, T.: Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis. In: Proceedings of EMNLP 2006, pp. 355–363 (2006)
8. Qiu, G., Liu, B., Bu, J., Chen, C.: Expanding Domain Sentiment Lexicon through Double Propagation. In: Proceedings of IJCAI 2009 (2009)
9. Kim, S.-M., Hovy, E.: Determining the Sentiment of Opinions. In: Proceedings of COLING 2004 (2004)
10. Esuli, Sebastiani, F.: Determining the Semantic Orientation of Terms through Gloss Classification. In: Proceedings of CIKM 2005, pp. 617–624 (2005)
11. Andreevskaia, A., Bergler, S.: Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In: Proceedings of EACL 2006 (2006)
12. Takamura, H., Inui, T., Okumura, M.: Extracting Semantic Orientations of Phrases from Dictionary. In: Proceedings of HLT-NAACL 2007, pp. 292–299 (2007)

13. Dragut, E.C., Yu, C., Sistla, P., Meng, W.: Construction of a sentimental word dictionary. In: Proceedings of CIKM 2010, pp. 1761–1764 (2010)
14. Zagibalov, T., Carroll, J.: Unsupervised Classification of Sentiment and Objectivity in Chinese Text. In: Proceedings of IJCNLP 2008 (2008)
15. Riloff, E., Wiebe, J., Wilson, T.: Learning subjective nouns using extraction pattern bootstrapping. In: Proceedings of CoNLL 2003, pp. 25–32 (2003)
16. Zhang, L., Liu, B.: Identifying Noun Product Features that Imply Opinions. In: Proceedings of ACL 2011, pp. 575–580 (2011)
17. Xu, Y., Liu, J., Seneff, S.: Mandarin Language Understanding in Dialogue Context. In: Proceedings of International Symposium on Chinese Spoken Language Processing (2008)

Personalized Fitting with Deviation Adjustment Based on Support Vector Regression for Recommendation

Weimin Li[1,2], Mengke Yao[1,2], and Qun Jin[3]

[1] School of Computer Engineering and Technology, Shanghai University, Shanghai, China
[2] Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai, China
[3] Colleage of Information Engineering, China Jiliang University, Hangzhou, China, and Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
{wmli,dream0708}@shu.edu.cn, jin@acm.org

Abstract. Almost all of the existing Collaborative Filtering (CF) methods rely only on the rating data while ignoring some important implicit information in non-rating properties for users and items, which have a significant impact on the preference. In this study, considering that the average rating of users and items have a certain stability, we firstly propose a personalized fitting pattern to predict missing ratings based on the trusty score set, which combines both the user-based CF and item-based CF. In order to further reduce the prediction error, we use the non-rating attributes, such as a user's age, gender and occupation, and an item's release date and price. Moreover, we present the deviation adjustment method based on the Support Vector Regression (SVR). Experiment results show that our proposed algorithms can increase the accuracy of recommendation versus the traditional CF.

Keywords: collaborative filtering, recommendation, personalized fitting, deviation adjustment.

1 Introduction

With the rapidly growing of information available on the Internet, people have to spend much more time selecting useful information. To solve the information overload problem, recommender systems have emerged. In recent years, recommender systems have widely used in e-commerce and social network to supply users with personalized information. Collaborative filtering is one of the most successful techniques for its simplicity and efficiency, and it is a good complementary technology to the content-based filtering. Its key process is to find similar users for the target user or similar items for the predicted item. However, there still exist some inherent problems to be addressed and solved, such as accuracy, data sparsity, cold start and scalability.

In order to improve the quality of the recommendation, various improved approaches, such as SVD [1], Biparite network [2] and Random walk [3], were introduced to collaborative filtering. Though the average ratings of users and items have shown a certain stability in a certain time period, these approaches, almost the same as the traditional similarity-based collaborative filtering, only consider the rating data, while ignoring many latent users and items features. For example, in the MovieLens

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_25, © Springer-Verlag Berlin Heidelberg 2014

datasets we can easily find that most students would prefer to the fantasy movies, and the popularity of comedy movies far surpasses drama.

In this paper, we make efforts to exploit a new score prediction method. In Section 2, we presents our design of a personalized fitting pattern by a using training set which comes from the trusty score set with regard to the target user and the target item, and we use the non-rating features which include both the user's and item's features to further lower the residual error. In Section 3, we propose the deviation adjustment mechanism by the Support Vector Regression (SVR) for the uncertain and nonlinear characteristic data. In Section 4, we show our experimental results to demonstrate that our proposed approach is more effective than both the traditional user-based CF and item-based CF.

2 Personalized Fitting

In this section, we define the similarity and trusty score set, and then outline the Personalized Fitting (PF) framework.

Let $\mathrm{U} = \{u_1, u_2, u_3 \dots, u_M\}$ be the set of users, and $\mathrm{S} = \{s_1, s_2, s_3 \dots, s_N\}$ be the set of items in the recommender system. We assume $r_{m,n}$ is the rating given to item $s_n \in S$ by the user $u_m \in U$. The history scoring record is presented by a $\mathrm{M} \times \mathrm{N}$ matrix .

There are many methods used to compute the similarity between the users and items in the collaborative recommender systems. The most popular methods are Cosin-based and Person Correlation Coefficient [4], which are all based on the rating matrix. The Person Correlation coefficient similarity can be given as follows.

$$sim(u_m, u_n) = \frac{|\sum_{s_i \in R_m \cap R_n} (r_{m,i} - \overline{u_m})(r_{n,i} - \overline{u_n})|}{\sqrt{\sum_{s_i \in R_m \cap R_n} (r_{m,i} - \overline{u_m})^2} \sqrt{\sum_{s_i \in R_m \cap R_n} (r_{n,i} - \overline{u_n})^2}} \tag{1}$$

where $R_m(R_n)$ is the set of records rated by $u_m(u_n)$. The correlation $\mathrm{corr}(u_m, u_n)$ between u_m and u_n is computed on the records $C_{m,n} = R_m \cap R_n$ rated by both u_m and u_n, and $\overline{u_m}$ and $\overline{u_n}$ indicates the average scores of u_m and u_n on all records of R_m and $\mathrm{R_n}$ respectively. We can also calculate the similarity between items s_m and s_n using the same principle as the user's similarity. All the similarity degree in this paper uses this Pearson Correlation Coefficient.

Similar to the similarity-based collaborative filtering method, our main purpose is also to predict the rating $\hat{r}_{m,n}$ of the target user $u_m \in U$ for the target item $s_n \in S$ that he/she has not known yet. But our predicting method is based on the Trusty Score set.

Definition 1. Trusty Score set (TS) of a target user u_m and a target item s_n are those rating records which belong to the existing ratings of Trusty Users set (TU) to the Trusty Items set (TI).

$$TS(u_m, s_n, k_u, k_s) = \{(u_i, s_j) | u_i \epsilon TU(u_m, k_u) \cap s_j \in TI(s_n, k_s) \cap r_{i,j} \neq \emptyset\} \tag{2}$$

where k_u and k_s are the trusty user size and trusty item size regulation parameters, respectively. $\mathrm{TU}(u_m, k_u)$ is the k_u nearest neighborhoods users of u_m , and $\mathrm{TI}(s_n, k_s)$ is the k_s nearest neighborhoods items of s_n. The Trusty Score set is a reliable training set in predicting rating since it takes the advantages of both the user-based and item-based collaborative filtering approaches.

To predict rating $\hat{r}_{m,n}$, there exists a linear relationship between the rating $r_{m,n}$ and the average rating of both u_m and s_n. The Trusty Score set $\mathrm{TI}(u_m, s_n, k_u, k_s)$ can be easily obtained when the adjusting parameters k_u and k_s are given. Here we suppose a tuple $(u_i, s_j) \in \mathrm{TI}(u_m, s_n, k_u, k_s)$, then we define the Personalized Fitting triples as $\delta_k(\overline{u_i}, \overline{s_j}, r_{i,j})$where $\overline{u_i} = (\frac{1}{|R_i|}) \sum_{s_j \in R_i} r_{i,j}$ and $R_i = \{s_j \in S | r_{i,j} \neq \emptyset\}$;$\overline{s_j} = \left(\frac{1}{|T_j|}\right) \sum_{u_i \in R_j} r_{i,j}$ and $T_j = \{u_j \in U | r_{i,j} \neq \emptyset\}$ and $0 < k \leq |\mathrm{TI}(u_m, s_n, k_u, k_s)|$. In order to simplify the later descriptions, we generalize the Personalized Fitting triples as $\delta_k(x_k, y_k, z_k)$.

The adjusting parameters λ_m and μ_n can be obtained by means of minimizing the following loss function.

$$Los(\lambda_m, \mu_n) = \sum_k (\lambda_m x_k + \mu_n y_k - z_k)^2$$

In general, the least square and gradient descent [5] can be used to minimize Formula (4), but the gradient descent method can get higher precision in the shortest time.

$$\begin{cases} \frac{\partial_{Los}}{\partial_{\lambda_m}} = 2\sum(\lambda_m x_k + \mu_n y_k - z_k)x_k \\ \frac{\partial_{Los}}{\partial_{\mu_n}} = 2\sum(\lambda_m x_k + \mu_n y_k - z_k)y_k \end{cases} \tag{4}$$

In this paper, firstly, we take the derivatives with respect to parameter λ_m and μ_n. Then, according to the gradient descent method, we should update the parameters along the gradienct descent direction. Therefore, the recursion formulas can be given as follows.

$$\begin{cases} \lambda_m = \lambda_m - \theta \frac{\partial Los}{\partial \lambda_m} \\ \mu_n = \mu_n - \theta \frac{\partial Los}{\partial \mu_m} \end{cases} \tag{5}$$

where the learning rate θ is assigned 0.001 generally, and parameters λ_m and μ_n are obtained by the gradient descent method. In this study, predicting the rating of the target user u_m to the item s_n can be expressed as Equation (6).

$$\hat{r}_{m,n} = \lambda_m \overline{u_m} + \mu_n \overline{s_n} \tag{6}$$

3 Deviation Adjustment by Support Vector Regression

In the previous section, we consider the rating $r_{m;n}$ only depends on the average rating $\overline{u_m}$ and $\overline{s_n}$. When we use the linear fitting model to describe their relationship, the predicted rating $\hat{r}_{m,n}$ can be obtained by the traditional CF or our proposed PF (Experiments using Personalized Fitting have shown better accuracy compared with the traditional similarity-based CF in Section 4). However, some other non-rating factors (such as a user's age, gender, occupation and an item's category, brand, ect.) are also important to affect $\hat{r}_{m,n}$. We assume a certain relationship exists between the residual $(r_{m,n} - \hat{r}_{m,n})$ and those non-rating factors. We propose a deviation adjustment method based on Support Vector Machine (SVM) to further improve the rating prediction accuracy.

Support Vector Machine was proposed by V.Vapnik in 1960s. It is a universal machine learning algorithm based on solid statistical theory foundation. SVM learning algorithms are based on the structural risk minimization, which is different from the empirical error minimization used in the traditional machine learning algorithms. What's more, SVM has shown its great advantage in small sample learning, nonlinear classification and poor generalization ability [6]. Moreover, V.Vapnik expanded SVM to regression forecasting by adding ε-insensitive loss function, and built the Support Vector Regression (SVR) theory. The essence of SVR is the convex quadratic optimization problem. Its discriminant function is given as follows.

$$\mathrm{f(x)} = \sum_{i=1}^{t}(a_i^* - a_i)K(x, x_i) + b \tag{7}$$

where $\mathrm{K}(\mathrm{x}, x_i)$ is the kernel function. Selecting the kernel function is the core step of SVR in solving the nonlinear regression problems. The basic idea is to transfer the original space into a new space $\Phi(\mathrm{x})$ by the kernel function. The new space is linearly separable. We only need to ensure the function like $\mathrm{K}(\mathrm{u}, \mathrm{v}) =< \Phi(\mathrm{u}), \Phi(\mathrm{v}) >$ because it only uses the dot product in our new training model.

In this model, the residual $(r_{m,n} - \hat{r}_{m,n})$ comes from the user features P_m and the item features Q_n. Here, we selected the user's feature to include: gender (P_u^1), age (P_u^2), and occupation (P_u^3) , and the item's feature to include: released year (Q_s^1) and genre (Q_s^2). Table 1 illustrates a simple example.

Table 1. A simple example training data for our model

u_m	s_n	deviation	P_u^1	P_u^2	P_u^3	Q_s^1	Q_s^2
1	15	0.35256	0	1	20	2	9
45	157	-0.24610	0	1	15	3	6
108	50	0.54350	1	2	4	1	4
204	123	1.23811	0	3	11	0	13
335	1001	-1.47634	1	1	8	2	4
….	….	….	…	…	…	…	…

A stable SVR model can handle the user features and the item features to a deviation which can adjust collaborative recommender methods.

4 Experiments

Our experiments were performed on a real and classical movie database MovieLens (http://www.movielens.umn.edu). The main datasets include 100,000 ratings from

943 users who reviewed 1,682 movies. The biggest advantage of MovieLens dataset is that we can easily extract non-rating features for our work. In our experiments, we have extracted important the user features which contain gender, occupation and age, the movie features which contain categories and release year.

In our work, we merely consider the precision as the only evaluation criterion to compare our method with the user-base CF and item-based CF. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are the most widely used indicators in collaborative filtering. MAE and RMSE are defined as follows.

$$\text{MAE} = \frac{\sum_{i=1}^{N}|r_{m,n}-\hat{r}_{m,n}|}{N} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}|r_{m,n}-\hat{r}_{m,n}|^2}{N}} \tag{9}$$

where $r_{m,n}$ is the actual rating that user u_m gave to item s_n in the testing set, while $\hat{r}_{m,n}$ is the corresponding prediction rating calculated by certain methods using the training set. N stands for the number of testing records. Therefore, the smaller MAE and RMSE are, the better prediction quality of related method is.

In this paper, in order to validate the effectiveness of our PF algorithm, we compare our PF method with the traditional collaborative filtering methods, including the user-base CF and item-based CF. In our PF algorithm, the trusty users adjusting parameter k_u and trusty items adjusting parameter k_s have a great influence on the experiment results. In order to intuitively reveal experiment results, we let $k_u = k_s = k$, where the k is the number of the nearest neighborhood (users or items) used in the traditional collaborative filtering. The experiment results are shown in Fig. 1.

As Fig.1 shown, our PF algorithm obviously has lower MAE and RMSE than the user-based CF and item-base CF. MAE decreased with the neighborhood size growing. However, when the neighborhood size reaches to 80, we can see MAE becomes bigger due to the over-fitting problem.

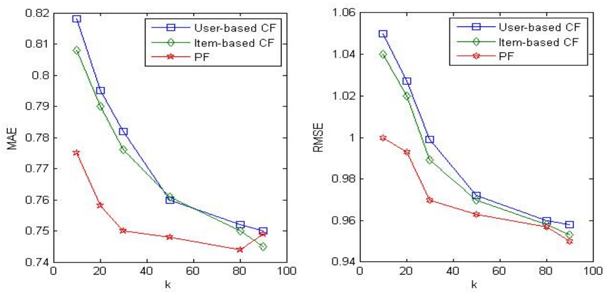


Fig. 1. Comparing our PF algorithm with CF by using MAE and RMSE curves

We selected Radica Basis Function (RBF) as the kernel function for the SVM model and invoked the SVM toolbox in MATLAB 2008 directly. In order to verify the effectiveness of our deviation adjustment mechanism, we selected $k_u = k_s = k = 80$, for k =80 is the best neighborhood size with regard to our PF algorithm. We also implemented the deviation adjustment by the BP neural network (BPNN). Detailed comparisons on MAE are shown in Table 2. From Table 2, we can see the deviation adjustment mechanism has indeed further lowered MAE with both the traditional collaborative filtering and our PF. SVM model is better than BPNN model for the reason that SVR could not fall into global optimal.

Table 2. Comparing different deviation adjustment models with SVR by MAE

Method	BASIC	BPNN	SVR
User-based CF	0.760	0.756	0.750
Item-based CF	0.761	0.755	0.749
PF	0.745	0.743	0.740

5 Conclusion

In this study, we have focused on how to further improve the accuracy of recommendation. We firstly proposed the Personalized Fitting method to predict the missing rating for the reason that the average rating of users and items shows great stability over a period of time. As we know, most traditional collaborative filtering methods only considered the rating data in the rating matrix. In our paper, we further presented a deviation adjustment mechanism based on SVR by using the non-rating features. The experiments revealed that the non-rating attributes contributed to reduce the prediction errors.

So far, our work didn't take timeliness into account. In the future, we should make efforts to consider this afctor and optimize our algorithms.

Acknowledgments. This work is supported by Shanghai Leading Academic Discipline Project (Project No. J50103), and Natural Science Foundation of Ningxia under Grant No. NZ12212.

References

1. Wu, M.: Collaborative Filtering via Ensembles of Matrix Factorizations. In: Proc. KDD Cup and Workshops, pp. 43–47 (2007)
2. Zhou, T., Ren, J., Medo, M., Zhang, Y.C.: Bipartite network projection and personal Recommendation. Physical Review E 76(4), 046115 (2007)
3. Fouss, F., Pirotte, A., Renders, J.M., Saerens, M.: Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaboratvie Recommendation. IEEE Trans. on Knowledge and Data Eng. 19(3), 355–369 (2007)
4. Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Trans. Knowledge and Data Eng. 17(6), 734–749 (2005)
5. Bell, R., Koren, Y.: Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In: Proc. IEEE International Conference on Data Mining (ICDM 2007) (2007)
6. Xiao, X.F., Xu, L.H., Zhu, Y.: Short-term Traffic Flow Prediction Based on SVM. Journal of Guangxi Normal Univeristy: Natural Science Edition 30(4), 13–17 (2012)

A Quick and Effective Method for Ranking Authors in Academic Social Network

Longjie Li[1], Xiaoping Wang[1], Qibin Zhang[2], Peng Lei[1], Min Ma[1], and Xiaoyun Chen[1,*]

[1] School of Information Science & Engineering, Lanzhou University, China
[2] Gansu Computing Center, Lanzhou, China
{ljli,chenxy}@lzu.edu.cn

Abstract. Effectively and efficiently ranking authors is a significant work in academic social network analysis. This paper investigates this problem and proposes a feasible method, QRank. First, QRank computes the initial rank score of an author based on the qualities of papers as well as the contributions of the author to those papers. In this step, QRank emphasizes the contribution of first author of a paper and can deal with the situation that only one author presents on a paper. Then, QRank refines the rank score of each author by the mutual influence between authors. Experimental results show QRank is the fast method and obtains better ranking results than the compared methods.

Keywords: QRank, author rank, paper quality, author contribution.

1 Introduction

Recently, social network analysis [1] has developed as a popular research area and attracted a lot of interest of researchers from different scientific fields. As a significant branch of social network, academic social network has been studied in many research works [2–9]. In this aspect, one important task is to identify the status of the individual, such as Rising Star [5, 6], Expert [10, 11], Bole [7], etc. Behind these work, the crucial thing is to rank authors in certain research domains. In this paper, we aim to present a feasible method to quickly and effectively rank authors.

PageRank [12] is increasingly widespread in social network analysis [8, 13–15]. Although it made a great success in ranking web pages, PageRank is not adequately competent for author rank problem. In [4], *AuthorRank* is proposed, which takes both frequency of co-authorship and total number of co-authors on papers into the weight. Furthermore, *PubRank* [5] joins the qualities of papers and the mutual influence between authors. PubRank computes the quality score of a paper based on the rank of the venue where it was published. However, the rank system used in PubRank is coarse and there is no a unique criterion to determine the level of each venue. To overcome the limitations of PubRank, *StarRank* [6] is proposed. It considers the contribution of an author in terms of the present order of the author on papers as well as

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_26, © Springer-Verlag Berlin Heidelberg 2014

the dynamic rank scores of publication venues. The dynamic score is computed using the entropy of the venue. StarRank thinks the venue which only accepts papers on specific research areas are of high level and hence has less entropy, and vice versa. But the entropy of a venue cannot really reflect the quality of the venue. A vital issue of StarRank is the equation in which the quality score of a paper is computed contrary with its intuition. They think that the less the entropy of a venue is, the better the quality of a paper is. However, the result acts counter to their wishes.

Besides the aforementioned issues, the introduced methods have two common blemishes. One is the time consuming and another is the situation that a paper only has one author. To solve these problems, we propose a new method, QRank (Quickly Rank), which can quickly and effectively rank authors. QRank includes two steps. The first step calculates the initial rank scores based on the qualities of papers as well as the present orders of authors on papers. To measure the quality of a paper, we use the quality score of the publication venue. In addition, we separate the first author and other authors by setting different weights. The second step refines the rank scores by taking the mutual influence between authors into account.

2 Related Work

To rank authors in academic social network, PageRank can be directly employed. And some other methods, AuthorRank, PubRank and StarRank, improve PageRank by combining some factors respectively. AuthorRank takes both the co-authorship frequency and number of coauthors on papers into the weight between two authors.

PubRank thinks ranking authors is associated with: 1) the mutual influence between authors; 2) the number of working papers; 3) the quality of published papers.

StarRank attempts to improve the PubRank in two aspects. One is the mutual influence between authors. Another is the qualities of papers. It defines the contribution of an author to a paper as the reciprocal of the author's present order on the paper and then defines the mutual influence between A_i and A_j as $CW(i,j) = (\sum c_i + \sum c_j) / \sum pc_j$. Here, $\sum pc_j$ is the total contributions of A_j in all papers, $\sum c_i$ is the contributions of A_i in co-authored papers of A_i and A_j. The average paper quality score of an author is computed based on the entropies of venues in which whose papers are published.

3 QRank

In order to rank authors in academic social network, two things that should be paid more attention are speed and accuracy. The QRank is a feasible method which can quickly and effectively rank authors in academic social network. The main ideas of QRank are listed in the followings: (1) To measure the quality of one paper we take the quality score of the venue that the paper was published. (2) Distinguish the contribution of each author on a paper, especially, the contribution of the first author. (3) Compute the initial rank score of an author based on the qualities of papers of this author and then refine the author's rank score based on the influence of other authors on this author. Based on these ideas, we divide QRank into two steps and describe the procedure at follows.

Step 1: We compute the initial rank scores of authors based on the qualities of papers. When calculating the author's score we give the first author and other author different weights. For other author, we also consider the present orders on papers, since the order can reflect the contribution of that author. Hence, we define the initial rank score of author A_i in Eq. 1.

$$IR(A_i) = \alpha \sum_j \lambda(p_j) + (1-\alpha) \sum_k \lambda(q_k) \frac{1}{o_k} \tag{1}$$

where p_j (q_k) is the j^{th} (k^{th})paper on which author A_i is (is not) the first author. $\lambda(p_j)$ ($\lambda(q_k)$) is the quality score of paper p_j (q_k). $\lambda(p_j)$ adopts the venue's quality score in which p_j was published. o_k is the present order of A_i on q_k. $\alpha \in (0,1)$ is a weight to adjust the importance of first authors.

Step 2: We refine the rank scores of authors using the co-authorship between them. So in this step we take the mutual influence into consideration to refine rank scores. Different with PubRank, AuthorRank, etc, QRank only passes the influences between co-authors once. Then we formally give the computation of QRank.

$$QRank(A_i) = IR(A_i) + \beta \sum_j IR(A_j) \frac{w_{i,j}}{\sum_k w_{j,k}} \tag{2}$$

where β ($0 < \beta \le 1$) is constant which is used to adjust the influence of the second part. $w_{i,j}$ is the number of co-authored papers between authors A_i and A_j. Therefore, $w_{i,j} / \sum_k w_{j,k}$ represents the ratio of the number of co-authored papers between A_i and A_j to the total number of papers that A_j participates.

Since QRank is a non-iterative algorithm, compared with other iterative methods, such as AuthorRank, PubRank, etc, it is very fast. And besides this, the effectiveness of QRank is guaranteed by our ideas mentioned above.

Lemma 1. *The time complexity of QRank is* $O(n)$ *and* n *is authors' number.*

Proof. Suppose k is the average number of papers that an author participates in and l is the average number of co-authors of an author. Clearly, the time complexities of step 1 and 2 are $O(kn)$, $O(\ln)$, respectively. Actually, $k, l \ll n$ (shown in Section 4), hence, the time complexity of QRank is $O((k+l)n) = O(n)$.

4 Experiments

In this section, we experimentally evaluate QRank over DBLP bibliography data [16], and compare it with PageRank and AuthorRank.

***Dataset and Parameter Settings*:** Two datasets, DM&DB and AI&PR, are extracted from DBLP. These datasets are derived from 10-year publications (2003-2012) in some major conferences in the fields of Data Mining, Database and Artificial

Table 1. Conference list

Conf.	Qualit	Conf.	Qualit	Conf.	Qualit	Conf.	Qualit
ICDE	65	VLDB	80	AAAI	44	ECAI	25
KDD	73	SIGMOD	88	CVPR	104	ECCV	57
WSDM	19	EDBT	27	ICCV	78	ICAPS	28
CIKM	43	PODS	48	ICML	70	ICCBR	16
ICDM	43	ICDT	27	IJCAI	5729	COLING	22
SDM	39			COLT	29	KR	24
PKDD	27			NIPS	74	UAI	36
PAKDD	23			ACL	56	AAMAS	48
DASFAA	18			EMNLP	33		
DM&DB				AI&PR			

Intelligence, Pattern Recognition, respectively. In Tab. 1, we list the selected conferences as well as the quality score of each conference. The quality score adopts the last 10-year field rating of each conference from Microsoft Academic Search [17]. Tab. 2 lists the statistics of the two datasets. The third and forth columns are the average papers and the average co-authors of an author. From Tab. 2, we clearly know that $k, l \ll n$ (**Lemma 1.**)

Table 2. Statistics on two datasets

Dataset	authors	papers	avg papers	avg co-authors
DM&DB	15293	11692	2.63	5.96
AI&PR	23000	20545	2.76	5.16

In our experiments, α and β are set to 0.65. The reasonable values of α and β are [0.6, 0.7] and [0.5, 1], respectively.

Experimental Results: In Tab. 3 and 4, we list the top 20 authors ranked by PageRank, AuthorRank and QRank from DM&DB and AI&PR, respectively. However, it is difficult to evaluate the results of each methods because there is no a unique criterion and different criterion may take the different estimated thoughts.

The H-index attempts to measure the influence of the publications of an author. Its value is the largest integer H such that H publications of an author have at least H citations. Although H-index only considers the citations of papers, it can be an auxiliary to some extent. Besides H-index, the average citation of papers is also employed as a reference. In Tab. 5, the mean H-index and mean average citations of top 20 authors ranked by the 3 methods in two datasets are illustrated. We obtain the H-indexes and citations of authors from Arnetminer [18]. From Tab. 5, we can see that QRank is the best among the 3 methods.

Running Time Comparison of QRank with Other Methods: We evaluate the running time of QRank compared with PageRank, AuthorRank. We extract five subsets from

DM&DB, AI&PR according years, respectively. The details of these subsets are listed in Tab. 6. Fig. 1 shows the running time of each compared method on these subsets. Clearly, QRank is the fastest method and is approximately linear. From Fig. 1, we conclude QRank is from several dozens to a few hundred times faster than PageRank and AuthorRank along with the increasing of number of authors.

Therefore, through the theory analysis and experimental studies, we claim that QRank is a quick and effective method to rank authors.

Table 3. Statistics of top 20 authors

Fields	Methods	Avg H-index	Avg Citations
DM&DB	PageRank	51.2	49.18
	AuthorRank	53.7	49.85
	QRank	**56**	**73.35**
AI&PR	PageRank	55.5	70.39
	AuthorRank	**59.75**	72.88
	QRank	58.85	**76.93**

Table 4. Years and number of authors

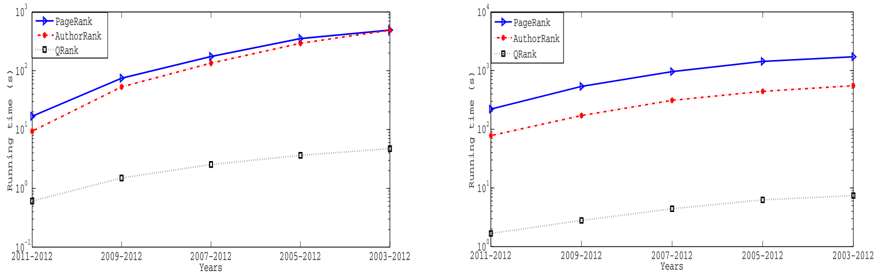
Years	Number of Authors	
	DM&DB	AI&PR
2011-2012	5557	9402
2009-2012	8924	13485
2007-2012	11521	17832
2005-2012	13671	20794
2003-2012	15293	23000

Table 5. Top 20 Authors from the DM&DB

PageRank	AuthorRank	QRank
Jian Pei	Philip S. Yu	Philip S. Yu
Philip S. Yu	Jiawei Han	Jiawei Han
Christos Faloutsos	Christos Faloutsos	Surajit Chaudhuri
Kian-Lee Tan	Jeffrey Xu Yu	Nick Koudas
Beng Chin Ooi	Jian Pei	Divesh Srivastava
Hui Xiong	Haixun Wang	Christos Faloutsos
Jeffrey Xu Yu	Divesh Srivastava	Jian Pei
Gerhard Weikum	Eamonn J. Keogh	Haixun Wang
Divesh Srivastava	Gerhard Weikum	Yufei Tao
Haixun Wang	Hans-Peter Kriegel	Hans-Peter Kriegel
Jiawei Han	E. A. Rundensteiner	Graham Cormode
Qiang Yang	Kian-Lee Tan	Jeffrey Xu Yu
A. K. H. Tung	Ming-Syan Chen	Charu C. Aggarwal
Jie Tang	Beng Chin Ooi	Rakesh Agrawal
E. A. Rundensteiner	Tao Li	Wenfei Fan
Eamonn J. Keogh	Hui Xiong	Sihem Amer-Yahia
Vanja Josifovski	Nick Koudas	Serge Abiteboul
Samuel Madden	A. K. H. Tung	Ruoming Jin
Aoying Zhou	S. Parthasarathy	Beng Chin Ooi
Ming-Syan Chen	ChengXiang Zhai	Xifeng Yan

Table 6. Top 20 Authors from the AI&PR

PageRank	AuthorRank	QRank
N. R. Jennings	N. R. Jennings	Xiaoou Tang
Andrew Y. Ng	Andrew Y. Ng	Shuicheng Yan
B. Scholkopf	Michael I. Jordan	Rong Jin
Shuicheng Yan	Milind Tambe	Kristen Grauman
Michael I. Jordan	B. Scholkopf	Michael I. Jordan
Luc J. Van Gool	Sarit Kraus	N. R. Jennings
Qiang Yang	Xiaoou Tang	Luc J. Van Gool
Sarit Kraus	Luc J. Van Gool	B. Scholkopf
Thomas S. Huang	Qiang Yang	Andrew Y. Ng
Milind Tambe	Mubarak Shah	Mubarak Shah
Larry S. Davis	Shuicheng Yan	Tuomas Sandholm
Mubarak Shah	Tuomas Sandholm	Eric P. Xing
Daphne Koller	Daphne Koller	A. Zisserman
Marc Pollefeys	Eric P. Xing	Thomas S. Huang
Nikos Paragios	Rong Jin	Stefano Soatto
Rong Jin	Peter Stone	Pascal Fua
Xiaoou Tang	Larry S. Davis	Larry S. Davis
Eric P. Xing	Thomas S. Huang	Vincent Conitzer
Pascal Fua	Pascal Fua	Marc Pollefeys
Csaba Szepesvari	Manuela M. Veloso	Daphne Koller



(a) DM&DB (b) AI&PR

Fig. 1. Running Time

5 Conclusions

This paper mainly focuses on the problem of quickly and effectively ranking authors in academic social network. Combining the qualities of papers and the influences between authors, we propose a new method, QRank. The basic idea of QRank is reasonable and the time complexity of QRank is $O(n)$. The experimental results show that QRank is superior to PageRank and AuthorRank.

References

1. Oliveira, M., Gama, J.: An overview of social network analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2(2), 99–115 (2012)
2. Jiang, X., Sun, X., Zhuge, H.: Graph-based algorithms for ranking researchers: not all swans are white! Scientometrics 96(3), 743–759 (2013)
3. Li, E.Y., Liao, C.H., Yen, H.R.: Co-authorship networks and research impact: A social capital perspective. Research Policy 42(9), 1515–1530 (2013)
4. Liu, X., Bollen, J., et al.: Co-authorship networks in the digital library research community. Information Processing & Management 41(6), 1462–1480 (2005)
5. Li, X.-L., Foo, C.S., Tew, K.L., Ng, S.-K.: Searching for rising stars in bibliography networks. In: Zhou, X., Yokota, H., Deng, K., Liu, Q. (eds.) DASFAA 2009. LNCS, vol. 5463, pp. 288–292. Springer, Heidelberg (2009)
6. Daud, A., Abbasi, R., Muhammad, F.: Finding rising stars in social networks. In: Meng, W., Feng, L., Bressan, S., Winiwarter, W., Song, W. (eds.) DASFAA 2013, Part I. LNCS, vol. 7825, pp. 13–24. Springer, Heidelberg (2013)
7. Yang, Z., Tang, J., et al.: Expert2bole: From expert finding to bole search. In: KDD 2009 (2009)
8. Yan, E., Ding, Y.: Applying centrality measures to impact analysis: A coauthorship network analysis. Journal of the American Society for Information Science and Technology 60(10), 2107–2118 (2009)
9. Lu, H., Feng, Y.: A measure of authors' centrality in co-authorship networks based on the distribution of collaborative relationships. Scientometrics 81(2), 499–511 (2009)

10. Zhang, J., Tang, J., Li, J.Z.: Expert finding in a social network. In: Kotagiri, R., Radha Krishna, P., Mohania, M., Nantajeewarawat, E. (eds.) DASFAA 2007. LNCS, vol. 4443, pp. 1066–1069. Springer, Heidelberg (2007)
11. Cai, Y., Chakravarthy, S.: Expertise ranking of users in QA community. In: Meng, W., Feng, L., Bressan, S., Winiwarter, W., Song, W. (eds.) DASFAA 2013, Part I. LNCS, vol. 7825, pp. 25–40. Springer, Heidelberg (2013)
12. Page, L., Brin, S., et al.: The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (1999)
13. Heidemann, J., Klier, M., Probst, F.: Identifying key users in online social networks: A pagerank based approach. In: ICIS 2010, p. 79 (2010)
14. Wang, R., Zhang, W., Deng, H., Wang, N., Miao, Q., Zhao, X.: Discover community leader in social network with pageRank. In: Tan, Y., Shi, Y., Mo, H. (eds.) ICSI 2013, Part II. LNCS, vol. 7929, pp. 154–162. Springer, Heidelberg (2013)
15. Bodendorf, F., Kaiser, C.: Detecting opinion leaders and trends in online social networks. In: ACM Workshop on Social Web Search and Mining, pp. 65–68 (2009)
16. `http://dblp.uni-trier.de/xml/`
17. `http://academic.research.microsoft.com/` (accessed on October 2013)
18. `http://arnetminer.org/` (accessed on October 2013)

Entity Information Retrieval System Based on EPC Network

Zhang Xinsong and Wang Dong

School of Software, Shanghai Jiaotong University, Shanghai, 200240, China
xszhang0320@gmail.com, wangdong@cs.sjtu.edu.cn

Abstract. With the further development of RFID and intelligent sensor technology, the Internet of Things develops from a fuzzy concept to the concreting technology. One of the typical applications of Internet of Things is EPC network. EPC (Electronic Product Code) network is an open-loop framework using EPC as unique identifier to organize entity, storing and sharing information through organizational boundaries. For the current EPC network, it has been achieved for an item level traceability of electronic information, but function of information retrieval is relatively weak. This paper deeply analyses the characteristics of EPC Discover Service and demand of entity information retrieval, and puts forward a rich discovery service to support entity information retrieval. Gaia is an entity information retrieval engine with good extension performance, flexible architecture and functional.

Keywords: EPC network, Information retrieval, similarity analysis.

1 Introduction

The Internet of Things [1] refers to uniquely identifiable objects and their virtual representations in an Internet-like structure. The term Internet of Things was proposed by Kevin Ashton in 1999. The concept of the Internet of Things first became popular through the Auto-ID Center at MIT and related market analysis publications. According to ABI Research more than 30 billion devices will be wirelessly connected to the Internet of Things by 2020. Under the guidance of Auto-Id laboratory, the development of Internet of Things has made great strides in recent years. A series of intelligent sensing technology like RFID apply to transportation and logistics, supply chain management, health care, intelligent environment and so on. Hence, it has very broad market and application prospect. EPC Network is a kind of typical network of Internet of things, presented by global standards organization GS1. It makes the supply chain information more accurate, instant and transparent by promoting the global application of EPC.

Traditional web information search cannot meet the physical requirements for structure data. However, with the development of the Internet of Things, the way we organize data makes it possible to complete real time entity search. Entity search engine technology has attracted industry attention of many researchers and academics.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_27, © Springer-Verlag Berlin Heidelberg 2014

2 Related Work

EPCglobal is an organization to promote worldwide EPC network specification and technology [2]. EPCglobal not only builds a complete EPC network infrastructure construction, but also focus on the future of Internet of Things and proposes strategic plan within the next decade. On the Auto-ID labs semiannual report 2012-2013 [3], GS1's CTO Steven Bratt summarized the technical and commercial challenges that EPC network will face over next ten years. And he organized ten research group under the jurisdiction of GS1 to report according to their research topic. The ten most concerned topics by GS1 are Antenna-based sensing, Meta-Materials, EPC calculus, supply chain security, standard scope expansion, visibility in a box, IOT GS1 digital architecture, digital product profiles, digital accounts and In-store promotion. We can conclude that EPC network will be expanded towards more diverse functions from the next decade plan of GS1. Expanded from B2B framework to B2B2C network framework is an inevitable trend. This paper aims to enrich EPC network entity information retrieval.

As the diversity and complexity of the application in Internet of Things, there is a lot of points designing an entity research system. Despite the large design space, many existing systems rely on a small set of fundamental techniques and strategies to implement entity discovery. Professor Kay Romer regards several technology as the most important ones in entity retrieval which are push and poll, publish and subscribe, mediators, inverted index, compression, models, scoring and ranking and Top-k query. Dyser system is a prototype of a search engine for real word, which is presented by research team lead by Professor Kay Romer [4]. Users can query the current state of the physical entity through Dyser system. In Dyser, there are two core elements, sensors and physical entities. Each sensor and physical entities will have a web resource description page. The system expresses all the resource description pages to HTML format and divides them into the sensor pages and physical pages. Despite basic query service, Dyser also provides three viable forecast model, which are APM (Aggregated prediction model), SPPM (single period prediction model) and MPPM (multi period prediction model). The models above are expansion to entity information retrieval.

3 Entity Retrieval System

Architecture

Entity information retrieval is an important extension of the network service. The objective of this study is to design a flexible, scalable, feature-rich entity information retrieval system based on EPC network infrastructure, which is called Gaia system. The overview of Gaia is described as Fig. 1.

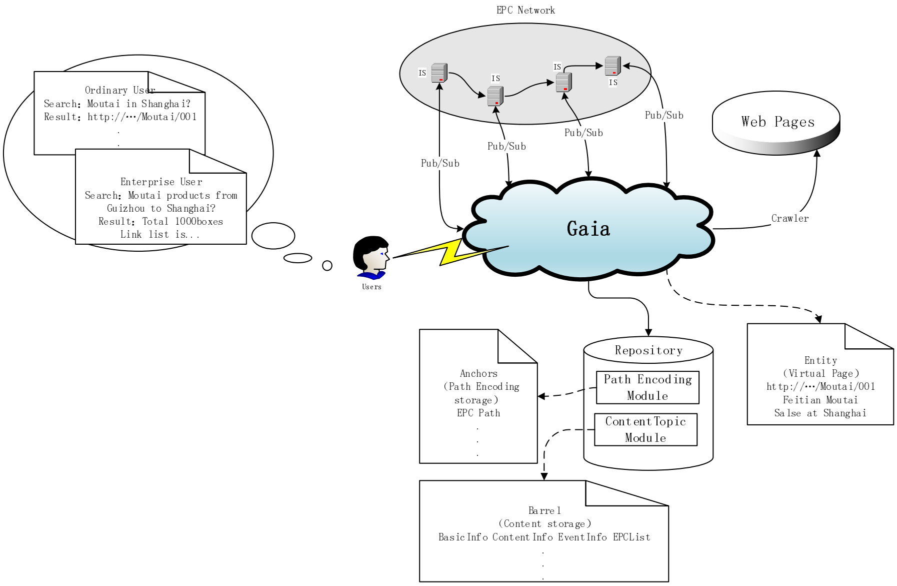


Fig. 1. Architecture of Gaia

Gaia system is based on EPC network infrastructure and draws from textual search engine Google's architecture [5]. Each entity in the system has a general description page. Most such objects produced by a registered EPCIS. When there is an event happened in EPCIS, the EPCIS will send a request for registering the event to Gaia system. Gaia will analysis the request and allocate storage space reasonable according to data visibility range. Two types of data are stored, path information and content topic. Two main functions of Gaia are statistical retrieval and content retrieval.

Statistical Retrieval

In the actual circulation of the product, the product stream flows through different information server. Therefore, it is difficulty to gain statistical information across nodes. In this paper, we come up with an efficient encoding strategy to store the path information, called prime path encoding strategy [6], to support statistical retrieval. Gaia assigns an appropriate prime number to each EPCIS. Path information is described with two major parameters which are PC (prime code) and NOC (Node Order Code). PC is the multiplication of the prime number of each node in the path. NOC defines the order of all the nodes in the path and it is calculated with the following equation set.

$$\begin{cases} NOC \equiv 1 \pmod{P(L_1)} \\ NOC \equiv 2 \pmod{P(L_2)} \\ \quad\vdots \\ NOC \equiv n \pmod{P(L_n)} \end{cases}$$

Based on PC and NOC, Gaia possess all the necessary information about path.

Traditional frequent path analysis algorithms, like Apriori, always have two obvious flaws, producing a large number of candidate set and scanning data set repeatedly. Gaia system organized data in a different way, so it is more convenient for frequent path mining. Our algorithm only need scan data set once and do not need to generate candidate sets. For example, for looking for frequent path whose length is k and whose minimum support is s, we present the pseudo code as Fig. 2.

Content Retrieval

Content topic information has five parts which are Basic Info, Path Info, Event Info, Evaluation Info and EPC Info. Basic information contains the basic property information of products. Path information is as mentioned last section. Event information described the state of the products and evaluation information is capture from users. The above information modules record and maintain entity information in four dimensions which are nature, state transformation, space-time transformation and objective evaluation.

```
Function findFrequencePath(Path list paths, min_support s , frequence k)
begin
      HashMap<Path , count> candidate;
      Set frePaths;
      for(Path p in paths){
         epcCount = 0;
         factorize p to L1,l2,...,Ln;
         if(n<k)
               continue;
         for(i=0;i<=n-k;i++){
               if(candidate contains (Li(i+1),L(i+2),...,L(i+k)){
                     epcCount = candidate.count+ p.epcNumber;
                     candidate.put(Li(i+1),L(i+2),...,L(i+k) , epcCount);
               }
               else{
                     candidate.put(Li(i+1),L(i+2),...,L(i+k) , epcCount);
                  }
         }
  }
      for(Path in candidate){
            if(count > s)
                  frePaths.add(Li(i+1),L(i+2),...,L(i+k));
      }
end
```

Fig. 2. Frequent path analysis algorithm

The modules are feature space of entity. EPC information records the EPCs which are accord with the special feature vector. Gaia system calculates the similarity between two content topics with modified cosine similarity model. Considering two content topics A and B, the standard cosine similarity is as the following equation.

$$similarity = \cos(\theta) = \frac{A \bullet B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

However, the importance of four information modules is different. We put forward a method to make it be possible to configure the weight of different modules.

$$similarity = \frac{\sum_i \alpha_i \times \cos(\theta_i)}{\sum_i \alpha_i}$$

We calculate cosine similarity on different modules in its own feature space, then get weighted average of the four modules. Modified cosine similarity ranges from 0-1. And when the similarity value is larger, the degree of similarity is higher.

Tracing System for Liquor Supply Chain – A Case Study

Tracing system for liquor supply chain of Shandong Jingzhi Corporation was started in January 2013. By the implementation of Jingzhi program, we want to build a manure system with Internet of Things and Entity Information Retrieval. It is a domestic demonstration in food traceability management. Gaia system has been used in the system. By combining the actual project in production environment, Gaia system shows the efficient entity retrieval and reassuring stability.

4 Conclusion

Entity information retrieval technology will play an important role in search engine in the future. Meanwhile, the development direction of the EPC network is from B2B to B2B2C model transformation. The paper is the first study of entity information retrieval technology within the EPC network architecture. It is a real promotion which makes EPC network closed to consumers. In this paper, we do a research in EPC network, but the method and thought can be generalized to all the area related with entity information retrieval.

References

1. International Telecommunication Union (ITU). ITU Internet Reports 2005: The Internet of Things. ITU Internet Reports series (2005)
2. EPCglobal. EPCglobalStandards Overview [EB/OL], http://www.gs1.org/
3. EPCglobal. SEMI-ANNUAL REPORT [EB/OL], http://www.gs1.no/
4. Romer, K., Ostermaier, B., Mattern, F., Fahrmair, M., Kellerer, W.: Real-Time Search for-Real-World Entities: A Survey. Proceedings of the IEEE (2010)
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks (1998)
6. Lee, C.-H., Chung, C.-W.: Efficient Storage Scheme and Query Processing for Supply Chain Management using RFID. In: SIGMOD - International Conference on Management of Data (2008)

Inferences for Multicriteria Decision Making

Sylvia Encheva

Stord/Haugesund University College, Bjørnsong 45, 5528 Haugesund, Norway
sbe@hsh.no

Abstract. A substantial part of the literature on decision making is often devoted to issues relating qualities and styles of a leader as an individual. Current technological and economic developments require both knowledge and skills within different fields and as a result a large number of decisions have to be made by incorporating opinions of several experts. Thus the need for new methods facilitating such types of decision making is increasing all the time. In this paper we discuss multi criteria decision making supported by fuzzy inferences.

Keywords: Inferences, Education, Knowledge representation.

1 Introduction

Decision making research has a considerable history with many notable contributions, [9]. In [11] decision making has been pointed to as the most critical responsibility a leader has. Further more it is stated that any roadmap that guides decision making should be a systematic one that shows leaders what to do and when to do it, what information he should use and how he should use it, [7] and [11]. According to [8] leaders who want to be effective ought to plumb the depths of their inner wisdom and contemplate the deepest values they hold for themselves and those they lead.

A substantial part of the literature on decision making is devoted to issues concerning qualities and styles of a leader as an individual. When however several persons and a number of alternatives are involved, one of the many difficult situations that can occur is that preferences related to f.ex. three alternatives a,b,c are presented as: a is preferred to b, b is preferred to c and a is preferred to c, [5]. In addition current technological and economic developments require both knowledge and skills within different fields and as a result decisions have to be made by incorporating several experts opinions. Thus the need for new methods facilitating cooperative decision making is increasing all the time.

The work presented in [2] provides a theoretically sound and at the same time practical approach for determining a set of skills' levels necessary for solving a set of problems. The method assumes judgements supplied by one expert. In many real life situations a decision maker (by this we mean a group of people authorised to make final decisions) is presented with opinions of a number of experts. This might happened because a problem is important and the outcome has serious consequences,

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_28, © Springer-Verlag Berlin Heidelberg 2014

or different parties are coming with their own expertise, or the data is collected in different time periods and/or different locations, f. ex. In this paper we consider situations where several experts are involved.

2 Preliminaries

A *context* is a triple (G, M, I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively [1], [3], [14]. A *concept* of the context (G, M, I) is defined to be a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $A^{'} = B$ and $B^{'} = A$. The *extent* of the concept (A, B) is A while its *intent* is B. A subset A of G is the extent of some concept if and only if $A^{''} = A$ in which case the unique concept of the which A is an extent is $(A, A^{'})$.

A structure $L = (L, \wedge, \vee, \otimes, \rightarrow, 0, 1)$ is called a complete residuated lattice, if (1) $(L, \wedge, \vee, 0, 1)$ is a complete lattice with the least element 0 and the greatest element 1; (2) $(L, \otimes, 1)$ is a commutative monoid; (3) $(\otimes, \rightarrow)$ is a residuated pair in L, i.e. $a \otimes b \leq c \Leftrightarrow a \leq b \rightarrow c, a, b, c \in L$. A complete residuated lattice $L = (L, \wedge, \vee, \otimes, \rightarrow, 0, 1)$ is called a complete involutive residuated lattice if it still satisfies $a = a^{cc}$ (the negation c is defined by $a^{c} = a \rightarrow 0$). For two fuzzy sets $\tilde{X}_1, \tilde{X}_2 \in L^U$, where L^U is the set of all L -fuzzy sets in U $S(\tilde{X}_1, \tilde{X}_2) = \bigwedge_{x \in U} (\tilde{X}_1(x) \rightarrow \tilde{X}_2(x))$. $S(\tilde{X}_1, \tilde{X}_2)$ is the subsethood degree and expresses the truth value of "each element of $\tilde{X}_1$ is an element of $\tilde{X}_2$", [4]. If U and A are two sets, then $\tilde{X}_i \leftrightarrow \tilde{B}_i$, ..., $\tilde{X}_n \leftrightarrow \tilde{B}_n$. If $\tilde{X}^{'}$ (or $\tilde{B}^{'}$) is given then calculate $\tilde{B}^{'}$ (or $\tilde{X}^{'}$) where $(\tilde{X}_i, \tilde{B}_i)(i = 1, ..., n)$ are the given fuzzy concepts, and $\tilde{X}^{'} \in L^U, \tilde{B}^{'} \in L^A$. The operator R_0 [12] is used as an implication. Another approach to work with data presented in form of intervals is presented in [6]. While the method allows graphical representation of an interval ordering it still does not allow multi criteria decision.

All lower case letters denote real numbers and the upper case letters denote the interval numbers or the closed intervals on $\mathbb{R}$ as in [10].

$A = [a_L, a_R] = \{a : a_L \leq a \leq a_R, a \in \mathbb{R}\}$, where a_L and a_R are the left and right limit of the interval A on the real line $\mathbb{R}$, respectively.

3 New Fuzzy Concepts

High education sector all over the world is experiencing difficulties due to less financial support and ever changing demands from industries among other things. Contents of existing courses have to be updated and new courses are introduced every year. Potential students need to know which kind of knowledge is expected from them to possess for a successful course completion while course designers have to cooperate in order to develop an applicable summary of requirements.

Suppose a new course is to be introduced at undergraduate level and be given at university level. A set of five new concepts $c_i, i = 1,...,5$ need knowledge from six supporting topics $t_j, i = 1,...,6$. When both concepts and supporting topics are sets of internally interrelated elements, additional work has to be done for developing applicable requirements. One of the very important issues to consider while developing a new course based on previous courses, is the level of similarities among topics and concepts included in previously given courses. Concepts and topics similarities are determined by course developers of the new course.

Concepts similarities:

$$
\begin{array}{cccccc}
c_1 & c_2 & c_3 & c_4 & c_5 & c_6 \\
0.2 & 0.4 & 0.1 & 0.5 & 0.6 & 0.3 \\
0.4 & 0.3 & 0.2 & 0.2 & 0.1 & 0.5 \\
0.5 & 0.2 & 0.7 & 0.5 & 0.8 & 0.7 \\
0.3 & 0.3 & 0.6 & 0.8 & 0.1 & 0.5
\end{array},
$$

Topics similarities:

$$
\begin{array}{cccccc}
t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\
0.6 & 0.2 & 0.5 & 0.7 & 0.4 & 0.4 \\
0.1 & 0.5 & 0.3 & 0.1 & 0.6 & 0.5 \\
0.2 & 0.8 & 0.1 & 0.3 & 0.1 & 0.4 \\
0.4 & 0.2 & 0.4 & 0.3 & 0.2 & 0.7
\end{array}.
$$

Applying R_0 operator and the approach presented in [2] we conclude that for a set of concepts $\tilde{C} = (0.4,0.3,0.5,0.2,0.4,0.6)$ the set of topics is

$$0.3,0.5,0.4,0.2,0.4,0.4 \leq \tilde{T} \leq 0.4,0.6,0.4,0.6,0.7,0.6$$

Suppose data in the form presented above originates from two sources (two departments are joint together and both parties have their own data). Both sources provide data in form of intervals. In this case corresponding intervals belonging to each source are to be compared. Thus $\tilde{T}$ will be represented as: $(0.3,0.4)$, $(0.5,0.6)$, $(0.4,0.4)$, $(0.2,0.6)$, $(0.4,0.7)$, $(0.4,0.6)$. The second source say $\tilde{T}_1$ will be presented with another six intervals: $(0.3,0.6)$, $(0.7,0.8)$, $(0.4,0.6)$, $(0.3,0.7)$, $(0.5,0.8)$, $(0.5,0.7)$. If there are no disjoint couples of such intervals, the decision maker can take a pessimistic decision by choosing the intersection of any two corresponding intervals or an optimistic decision by choosing the union of any two corresponding intervals. If there are disjoint couples of intervals the decision maker can a apply the acceptability index, [10].

For the sake of presentation simplicity we refer to three sources only. A larger number of sources can be handled in a similar way. Suppose a decision maker receives data provided in a form of intervals from three sources, denoted s_1, s_2 and s_3. If s_1 is better than both s_2 and s_3 when the acceptability index has been applied, the decision maker takes obviously s_1. Suppose s_1 and s_2 are equivalent but better than s_3 when the acceptability index has been applied. The decision maker can either take any of the s_1 or s_2, or involve an additional expert. Taking at random one of the two options (note that they might be more than two if a larger number of sources is involved) might not be quite up to the decision maker satisfaction, and working with additional expert might be not only costly and time consuming but simply impossible due to f. ex. unavailability of such expert.

4 Conclusion

In this paper we address the problem of calculating new fuzzy concepts using information about known fuzzy concepts. In particular we propose a solution for cases from high education where new courses are to be introduced and preliminary knowledge are a prerequisite. One of the difficult questions to be answered is what is the level of similarity between new concepts and preliminary knowledge. Answering such questions by applying fuzzy inferences can considerably simplify the related decision making processes.

References

1. Davey, B.A., Priestley, H.A.: Introduction to lattices and order. Cambridge University Press, Cambridge (2005)
2. Fan, S.-Q., Zhang, W.-X., Xu, W.: Fuzzy inference based on fuzzy concept lattice. Fuzzy Sets and Systems 157, 3177–3187 (2006)
3. Ganter, B., Wille, R.: Formal Concept Analysis. Springer (1999)

4. Goguen, J.: L-fuzzy sets. J. Math. Anal. Appl. 18, 145–174 (1967)
5. Klahr, D.A.: Computer Simulation of the Paradox of Voting. The American Political Science Review LX(2), 384–390 (1966)
6. Kaytoue, M., Assaghir, Z., Messai, N., Napoli, A.: Two Complementary Classification Methods for Designing a Concept Lattice from Interval Data. In: Link, S., Prade, H. (eds.) FoIKS 2010. LNCS, vol. 5956, pp. 345–362. Springer, Heidelberg (2010)
7. Kepner, C.H., Tregoe, B.B.: The Rational Manager: A Systematic Approach to Problem-solving and Decision making. McGraw-Hill, New York (1965)
8. Klenke, K.: The factor in leadership education, practice, and research. Journal of Education for Business 79(1), 56–60 (2003)
9. Nutt, P.: Making decision-making research matter: some issues and remedies. Management Research Review 34(1), 5–16 (2011)
10. Sengupta, A., Pal, T.K.: On comparing interval numbers. European Journal of Operational Research 127, 28–43 (2000)
11. Trauffer, H.C.V., Bekker, C., Mihai Bocarnea, M., Bruce, E., Winston, B.E.: Towards an understanding of discernment: a conceptual paper. Leadership & Organization Development Journal 31(2), 176–184 (2010)
12. Wang, G.: Non-Classical Mathematical Logic and Approximate Reasoning. Science Press, Beijing (2010)
13. Ward, M., Dilworth, R.P.: Residuated lattices. Trans. Amer. Math. Soc. 45, 335–354 (1939)
14. Wille, R.: Concept lattices and conceptual knowledge systems. Computers Math. Applications 23(6-9), 493–515 (1992)

Navigation Mechanism in Blended Context-Aware Ubiquitous Learning Environment

Chuang-Kai Chiou[1], Judy C.R. Tseng[1,*], and Tien-Yu Hsu[2]

[1] Department of Computer Science and Information Engineering,
Chung Hua University, Hsinchu, 300, Taiwan, ROC
{d09524003,judycrt}@chu.edu.tw
[2] Department of Operation, Visitor Service, Collection and Information Associate Curator,
National Museum of Natural Science, Taichung, Taiwan, ROC
dan@mail.nmns.edu.tw

Abstract. In recent years, navigation support problems have been discussed and investigated in ubiquitous learning environment. Several researches have proven that students can perform better when navigation supports are provided by learning systems. However, traditional ubiquitous learning environments suffer from some physical limitations. For example, the capacities of learning targets are limited and/or moving times for reaching learning targets are required. To address these limitations and create a more efficient ubiquitous learning environment, a novel learning framework, namely the blended ubiquitous learning environment, is proposed. A blended navigation algorithm, B-MONS, is also proposed for developing a navigation support mechanism which suits the new learning framework. Experimental results show that students learn in the blended navigation environment with the help of B-MONs get higher learning performance.

Keywords: ubiquitous learning, e-learning, blended learning, navigation support.

1 Introduction

Several researchers have demonstrated the benefits of context-aware ubiquitous learning in helping students to improve their problem-solving ability in the real world [1-5]. However, they suffer a common problem which does not occur in the traditional learning environment, that is, how to provide an appropriate learning path for students. In practical applications, usually a fixed learning path is provided for all of the students without considering personal or environmental situations. When the diagram shifts to a u-learning environment, the pre-set learning path does not work anymore due to the fact that the constraints in such an environment are quite different from those in e-learning or m-learning.

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_29, © Springer-Verlag Berlin Heidelberg 2014

Researchers have paid considerable attention to solving the navigation problems in authentic learning environments [6, 7]. Because there are more real-time criteria to be considered in the authentic world, the navigation problem of context-aware u-learning becomes even more difficult than that of a web-based learning environment. In our previous work[6], a greedy method based navigation algorithm was proposed and verified by a serial of experiments that it would improve students' learning performance in ubiquitous learning environment.

Although the navigation support problem has been discussed and investigated, there are some problems that have not yet been addressed. For example, when a u-learning environment is overcrowded, students need to spend a lot of time waiting for the capacities of learning targets to be released. Another problem is that when the learning targets are scattered in a large learning space, learners need to spend a lot time moving across learning targets. These problems cause low learning performance in ubiquitous learning environments.

To cope with these problems, a novel framework, namely the blended ubiquitous learning environment, is proposed. The new environment combines the advantages of both e-learning and u-learning. Moreover, a navigation algorithm, B-MONS, is also proposed to develop a navigation support mechanism which suits the new framework. A serial of experiments are conducted to verify the proposed navigation algorithm. The results show that students' learning performance are improved when they learn in the proposed blended learning environment with the help of the new proposed navigation algorithm.

2 Relative Works

The benefits of context-aware ubiquitous learning in helping students to improve their problem-solving ability have been demonstrated by many researches [1-5]. For example, Hwang et al. [8] proposed a knowledge engineering approach to develop a Mindtool for context-aware u-learning. The developed Mindtool has been applied to a learning activity of a natural science course in an elementary school. The experimental results depict that the approach not only promotes learning motivation, but also improves the learning achievements of individual students.

In order to improve learning efficiency, several studies have employed optimization techniques to find a better learning path. For example. Hwang et al. [9] proposed a decision-tree-oriented guidance mechanism for context-aware ubiquitous learning environment. They found that representing knowledge with a decision tree is more suitable than the approaches used in neural networks, Bayesian networks, or rule-based engines. In our previous work [6], we considered the contextual information of real-time situations to determine the navigation path for learning activities. An adaptive navigation algorithm call MONS (Maximized Objective Navigation Support algorithm) was proposed. In the MONS algorithm, an objective function is designed for determining the priority of learning targets. The objective of MONS is to guide students along an efficient learning path that takes less learning time and has more learning efficacy.

Although these navigation mechanisms have been proven that the work well in improving the learning efficiency of students, some situations are not considered in previous studies.. For example, when the u-learning environment is overcrowded and/or when the learning targets are scattered in a large learning space, traditional u-learning environments will not be able to provide efficient learning.

3 Blended Ubiquitous Learning Environment and Its Navigation Support

In order to improve the learning efficiency of u-learning environments, a novel framework of learning environment, blended ubiquitous learning environment, is proposed. A navigation algorithm, B-MONS, is also proposed to guide students along efficient learning paths in the b-learning environment.

3.1 The Blended Ubiquitous Learning Environment

In this paper, we have re-defined the original environment model of u-learning in that we have blended the u-learning environment with e-learning environment. In other words, the blended u-learning (b-learning) environment is composed of a physical layer (inherited from u-learning) and a cyber layer (inherited from e-learning). Learning targets in the cyber layer are projections of learning targets in the physical layer. An illustration of b-learning is shown as Fig. 1.

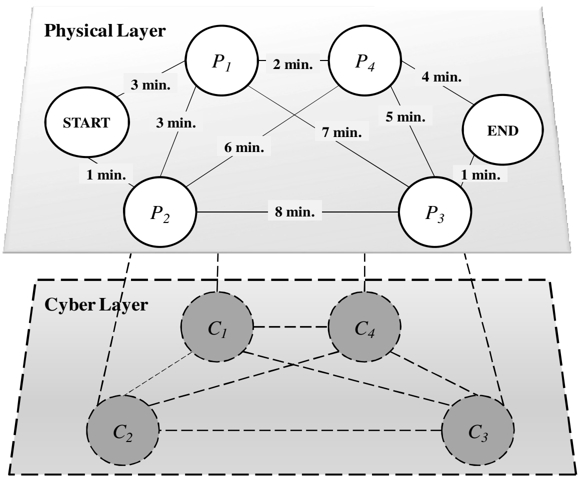


Fig. 1. An illustration of the blended ubiquitous learning environment

Unlike the original u-learning environment in which all learning activities are triggered by moving closer enough to the physical learning targets, b-learning environment allows learners to directly access the cyber learning materials from anywhere in the learning space. This property breaks down the capacity restriction of learning targets. When the physical learning targets in the b-learning environment are overcrowded, learners can transfer their learning to cyber learning targets. Similarly, when the distance between two learning targets is too far, learners can learn the cyber learning targets instead of wasting time walking to the physical learning targets. Hence, the blended framework inherits the advantages of both e-learning and u-learning.

3.2 Navigation Algorithm for the Blended Ubiquitous Learning Environment

In order to provide navigation guidance to learners, a maximized objective navigation support algorithm for the B-learning environment (B-MONS) is proposed. B-MONS is an extended version of MONS [6]. MONS only consider the physical learning target, but B-MONS consider both the physical and cyber learning targets. Instead of navigating students in a pure u-learning environment, B-MONS can navigate students in a B-learning environment to learn across physical layer and cyber layer.

When a learner is going to leave the current learning target D_i, the next learning target D_j will be determined by an objective function. All the non-visited physical and cyber learning targets will be determined by the evaluation function $Score\_X(D_i, D_j.)$ as shown in equation (1), and the solution with maximum profit in the current stage will be provided to the learner as D_j.

$$Score\_X(D_i, D_j.) = \left\{ \begin{array}{ll} \dfrac{G_j * ln(S - R_j + 1)}{MT_{i,j} + T_j} & if \;\; D_j \in PV \\ \alpha\left(\dfrac{G_j}{T} * \beta\right) & if \; D_j \in CV \end{array} \right\}$$
$$\text{where } \beta = \sum_{j=1}^{n} \frac{G_j * ln(S - R_j + 1)}{T_j} \Big/ \sum_{j=1}^{n} \frac{G_j}{T} \qquad (1)$$

In equation (1), n represents for the number of learning targets, $MT_{i,j}$ represents for expected time for moving from P_i to P_j, T_j represents for expected time for learning the target D_j. G_j represents for the expected learning profit of a learning target (C_j or P_j). It is a value ranging from 0 to 1. S represents for acceptable saturation ratio, which is the acceptable percentage of used capacity for learning targets. R_j represents for current saturation ratio of D_j and $R_j = M_j / PL_j$. α is the weight of the cyber learning targets. The value range of α is from 0 to 1. If α equals 0, the learners will not be guided to cyber layer. If α equals 1, the learners will have an equivalent chance of being guided to the cyber learning targets or the physical learning targets. β is a normalized coefficient which estimates the ratio of physical targets to cyber targets.

From equation (1), we can find that when there are too many learners in the B-learning environment or the moving time between the learning targets is too long, the evaluation score of the physical learning targets will be less than that of the cyber learning targets. In this situation, the learner will be guided to the cyber learning targets and learning the cyber learning material at the current location.

4 Performance Evaluation

In order to verify the feasibility of our framework and navigation algorithm, a simulation is conducted. A set of test data is executed in a simulation environment to evaluate the efficacy and immediacy of our algorithm. Two navigation algorithms, MONS and B-MONS are used to compare the average learning profit. The learning profit of a student is defined as the total learning profit of the learning targets that the student has visited. B-MONS guides learners to learn physical learning targets or cyber learning targets depending on the evaluation scores, but MONS only guides learners to learn physical learning targets. The influence of the number of learners on learning efficiency and the influence of the moving time between learning targets on learning efficiency were analyzed in the simulation experiment.

Because learning activities in an e-learning environment lack concrete experiences of observations and operations, it is assumed that the learning profit of cyber learning target would be less than that of physical learning targets. A reduce factor is applied on cyber learning targets to reflect this assumption. The reduce factor is set to 0.25 in our simulation.

The learning profit comparisons of MONS and B-MONS under different number of learner scales and learning targets scales are shown in Fig. 2.

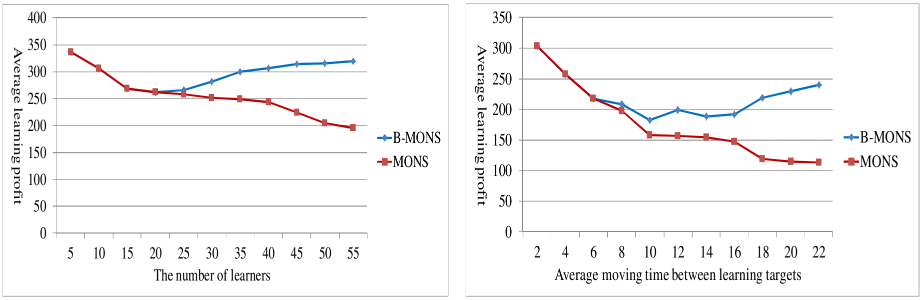


Fig. 2. Comparisons of B-MONS and MONS

From Fig. 2, we can find that B-MONS outperforms MONS. The low performance caused by over-crowded learning targets and long moving time between learning targets is improved. The simulation result therefore conforms to our assumption.

5 Conclusion and Future Works

In this paper, a novel framework, namely the blended ubiquitous learning environment, is proposed. Moreover, a navigation support mechanism which suits the new framework is also proposed and implemented. Such a new navigation mechanism overcome some limitations of the original u-learning environment. By allowing students to learn in cyber space, times wasted on waiting and moving when the learning environment is overcrowded or when the learning targets are scattered in a large learning space can be reduced. From the analytical results, we find that learning in

B-learning environment with the help of the B-MONS navigation algorithm results in higher learning efficiency and learning achievement than learning in original u-learning environment with the help of MONS.

The B-learning environment is a novel learning environment, there exists various research issues which can be further investigated. In the future, we will conduct some field experiments to verify the performance of our work. Moreover, we will also try to investigate the personalization issues in b-learning environment to learn more about the suitability of these learning environments, e-learning, u-learning and b-learning, for students with different learning style.

Acknowledgment. This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 102-2511-S-216-002-MY3.

References

1. Hsu, J.-M., Lai, Y.-S., Yu, P.-T.: U-plant: a RFID-based ubiquitous plant learning system for promoting self-regulation. International Journal of Internet Protocol Technology 6(1), 112–122 (2011)
2. Hwang, G.-J., et al.: A context-aware ubiquitous learning environment for conducting complex science experiments. Computers & Education 53(2), 402–413 (2009)
3. Yin, C., et al.: Supporting the acquisition of Japanese polite expressions in context-aware ubiquitous learning. International Journal of Mobile Learning and Organisation 4(2), 214–234 (2010)
4. Huang, Y.-M., et al.: A ubiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. Computers & Education 58(1), 273–282 (2012)
5. Tsai, P.-S., Tsai, C.-C., Hwang, G.-J.: Developing a survey for assessing preferences in constructivist context-aware ubiquitous learning environments. Journal of Computer Assisted Learning 28(3), 250–264 (2012)
6. Chiou, C.-K., et al.: An adaptive navigation support system for conducting context-aware ubiquitous learning in museums. Computers & Education 55(2), 834–845 (2010)
7. Hwang, G.-J., et al.: A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning. Computers & Education 54(2), 404–415 (2010)
8. Chu, H.-C., Hwang, G.-J., Tsai, C.-C.: A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning. Computers & Education 54(1), 289–297 (2010)
9. Hwang, G.-J., et al.: A Decision-Tree-Oriented Guidance Mechanism for Conducting Nature Science Observation Activities in a Context-Aware Ubiquitous Learning Environment. Educational Technology & Society 13(2), 53–64 (2010)

Personalized Itinerary Recommendation Based on User-Shared Trails

Hao Lu and Dong Wang

School of Software, Shanghai Jiao Tong University,
Shanghai, China, 200240

Abstract. Itinerary is a kind of abstraction of human life. However, with substantial effort involved, it is complicated for someone to plan an itinerary. Therefore, personalized and reasonable itinerary recommendation is nontrivial. With the development of social networks, people have shared their trails like photos, check-ins and text comments on locations. Historical itineraries can be extracted from these trails and make it possible to recommend new ones. Existing studies mainly focus on isolated locations. They firstly choose eligible locations and then construct new itineraries with the filtered locations and various contexts. However, the connections between locations have been ignored. In this paper, we propose the concept of transfer. We predict user preferences based on their actions on the transfers and link the transfers to itineraries. We evaluate our method on a foursquare check-in dataset and have achieved promising results.

Keywords: itinerary recommendation, context.

1 Introduction

As social networks mature, sharing on website has become an essential part in more and more peoples' daily lives, including sharing their location, photos, and comments on certain locations or events. These accumulated data can help us to determine users' location information at certain timestamps. This information are the basis of itinerary recommendation. While planning their itineraries, users will face two following problems: first, how to plan itineraries in perfect accordance with their own preferences; second, planning an itinerary involves selecting locations, determining the order to visit each location and other work as boring and time-consuming as the former two [1]. How to reduce the time and effort is quite necessary. Naturally, effective and reasonable itinerary recommendation is the answer to all these problems.

There has been some literature studying the issue of itinerary planning and recommendation. [2] and [3] discuss how to utilize geo-tagged photographs to make travel route recommendation. [4] studies the use of statistical model to exploit check-in data from social networks to build itineraries. [5] recommends social itineraries by learning from GPS trajectories uploaded by taxis. [1] is a research of interactive itinerary planning, in which each iteration will collect user feedback, adjust the itinerary

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_30, © Springer-Verlag Berlin Heidelberg 2014

according to the feedback and again collect further user feedback until terminated. These studies have focused on choosing discrete locations and link them together into itineraries. Apparently the implicit connections between locations have been ignored. In addition, while choosing locations, user preferences have only been taken into consideration in [2]. [1] has taken advantage of direct feedback manually provided by users instead of extracting user preferences from given data set. Other studies all select locations from a point of view of popularity to construct itineraries without much personalization.

In this paper, we propose the concept of transfer. Based on it, we use historical itineraries to make personalized recommendation. There are two aspects of personalization. Firstly, itineraries recommended should be consistent with their preferences. Secondly, it is the context of an itinerary that should also be personalized. Considering that a lot of places have obvious temporal attributes, i.e. people will get better experience to visit a place at a specific time or during a certain period of time. With different habits and preferences of different users, it is significant to identify the appropriate visit time to meet their implicit requirements.

2 Related Work

There have been some literatures studying the issue of itinerary recommendation. [5] uses the HITS algorithm to calculate the popularity of locations on the basis of a data-set of GPS coordinates uploaded by taxis. Then, Hyoseok et al. construct itineraries with time ratio and frequent patterns and sequences and sort the results. In [2], Kurashima et al. cluster the GPS metadata of photographs using mean-shift algorithm and propose a photographer behavior model trained by historical itineraries. With current location and time budget of the user, they recommend itineraries based on the behavior model to meet user preferences. [1] introduces explicit feedback of users into the planning process of itineraries. With binary POI feedback model, selection of POI in a new iteration is based on the user feedback in the previous one. Then, it continues collecting user feedback until the whole process terminates.

We find out that nearly no literature has achieved significant breakthrough in personalization of itinerary recommendation. The photographer behavior model proposed in [2] is a combination of Markov model and topic model, which can identify the user preferences for different locations to some degree. When it comes to selection of locations, most documents like [6] and [7] tend to choose popular ones. It is apparently far from enough to recommend itineraries only containing popular locations to users. In terms of itineraries, not only do we need to consider personalization in locations, but we also need to focus on personalization in various context. What's more, existing studies are all location-based. The proposed concept of transfer is more intuitive and more comprehensive due to taking the implicit connections between locations into account as well.

3 Methodology

3.1 User Preferences Extraction

With the rating matrix built from historical itineraries, we use matrix factorization to extract user preferences. Assume that there are a few latent factors which indicate user preferences, both users and transfers are related to the factors. If we have a rating matrix and value of each cell is the evaluation of a certain user's opinion on a specific factor, this matrix would be the representation of relationship between users and the latent factors. With another matrix for transfers and factors, we can evaluate the relationship between them. Finally, the multiplication of the two matrices would be the rating matrix M between users and transfers. That's why we can use factorize the rating matrix M to get the two factor matrixes.

The latent factors here might be the combination of multiple indicators which are both difficult to find out and hard to explain. With transfers as items, not only do the factors contain user preferences on locations, but also they reflect the probability of certain sequences of locations. For instance, user u might be fond of both location l_1 and l_2. If we only consider the interest, it's appropriate to recommend user u to go to l_2 after visiting l_1. However, people might seldom visit l_1 and l_2 consecutively due to the distance between them or the order that being visited or other reasons. These implicit reasons can be found in historical itineraries and therefore recorded in the relationship matrixes.

To factorize the rating matrix M, we use the common algorithm gradient descent. After a certain number of iterations, we can get the convergent factor matrixes. With the factor matrixes, we can also calculate the similarity of users and transfers. For example, we can find similar users by finding small difference between the corresponding vectors in user-factor matrix, which would be useful while calculating personalized context.

3.2 Context Calculation

As mentioned before, visit time of transfer $t = (l_1, l_2)$ is the timestamp that someone reaches location l_2. It's one of the most important contexts of an itinerary. Different locations have different temporal attributes. For instance, business hour of various shops and attractions could be different. People could only enjoy certain experience and groceries at specific times of a day. Particular landscape such as sunrise and sunset also has its own temporal attributes. Consequently, proper visit time for the end location of a certain transfer has a significant influence on the experience and the context of each transfer in the recommended itinerary affect the quality of the itinerary undoubtedly.

To determine the proper visit time of a transfer, [4] proposes a concept of temporal visiting distribution. However, there are some flaws in this concept. In most circumstances, people don't share their trails so frequent that rarely more than two records have same timestamp with same location. With same end location, transfers might vary in start locations as well. This kind of sparseness leads to a complete flat distribution with each circumstance occurs for only once, which is meaningless for calculating the context. So we propose a density-based visit distribution to better describe the context in historical data. The equation to calculate the probability of visit time t is as following.

$$\mathrm{P(t)} = \frac{\sum_{x\in(t-\varepsilon,t+\varepsilon)}(\varepsilon-|t-x|)^2}{\sum_{x\in(t-\varepsilon,t+\varepsilon)}1} \tag{1}$$

Furthermore, different people will prefer different visit time even for the same end location. For example, a distribution with multiple peaks like Figure 1 indicates that there are four small period of time that people tend to visit one place. Each person normally would visit this place at only one of the four timestamps. To take this kind of preference into consideration, we alter the density-based distribution into a weighted distribution with each circumstance multiplied by the similarity of that user of the record at that timestamp.

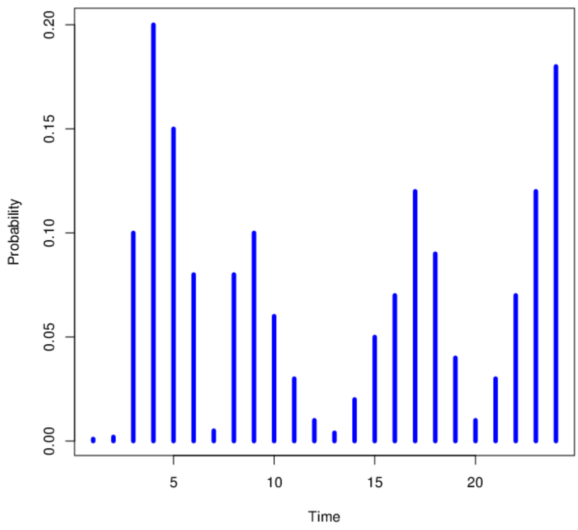


Fig. 1. Weighted distribution with multiple peaks

With this weighted density-based visit distribution (WDD), we use symmetric Kullback-Leibler Divergence [8] to calculate the fitness of a thin Gaussian distribution, which simulates the value we are going to set for the visit time. The smaller the divergence value is, the more proper the visit time is.

3.3 Recommended Itineraries Construction

An obvious thought to construct itineraries would be breadth first search (BFS) algorithm, which starts with the first location and keep adding new locations to the itinerary. Since transfers are all from historical itineraries, we have reasons to believe that moving by transfers which didn't exist in history are not proper in reality from a collective perspective. With this assumption, we need only to traverse all existing transfers which greatly scales down the searching space and reduces the complexity of the problem.

We first calculate the estimated visit time of current transfer by adding average duration to current timestamp. Then we evaluate the fitness of the estimated visit time at location l for user u. Finally, we update the itinerary and its total fitness value. When the loop is terminated, all possible itineraries have been constructed.

4 Evaluation

To evaluate the effectiveness of the method, it's intuitive to try finding out users' feedback on the recommended itineraries. However, we cannot conduct a survey to present recommended itineraries to users and then collect their feedback. In most cases, we use historical itineraries to evaluate as well.

According to our theory, the higher the fitness score is, the better the itinerary will be. Therefore, we decide to validate the fitness score to evaluate the quality of the itinerary. Concretely speaking, we choose one location in the itinerary as the start point and use the visit time from historical data as input. Then we calculate the fitness score of each transfer that already exists in history and check whether the one we choose has got the highest fitness score in turn. Take transfer $t = (l_1, l_2)$ for an example, suppose t_v is part of an itinerary of user u. The visit time of t in history is t_v. Then we can calculate the fitness score as $WP(t_v, u)r(t', u)$, while t' here is one of all possible transfers with l_1 as the start location. After the calculation, we check the ranking of the fitness score that location l_2 has got and make it the accuracy indicator.

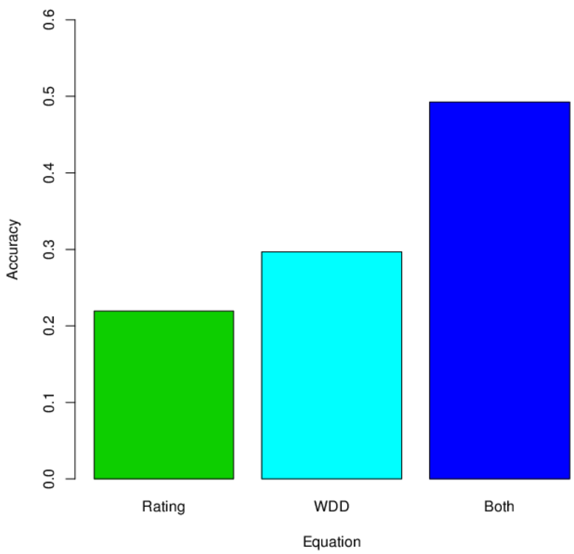


Fig. 2. Accuracy of different fitness equations

We also omit part of the fitness equation to make comparison. First, we ignore the user preferences so that the fitness score would be only weighted visit time distribution. Then, we ignore the context and regard the rating as its fitness score. Through the comparison, we can validate the effectiveness of each part of the fitness score. With the above steps, we can check the necessity of choosing the right transfer and calculate the proper context. And the result in Figure 2 shows that the accuracy is higher without any omission.

5 Conclusion

In this paper, we propose the concept of transfer and introduce the method based on transfer to recommend itineraries. We believe that the user preferences can be extracted from historical itineraries. We factorize rating matrix to reveal the relationship between users and transfers and use weighted density-based visit time distribution to calculate the proper context. Then we design an algorithm to construct itineraries. Through comparison in experiments, we validate the effectiveness of the method. Our major future work is to study recommending itineraries with hierarchies to better adapt to the granularity of the dataset. Besides, privacy and trustworthiness are definitely the major concern as well.

References

1. Basu Roy, S., Das, G., Amer-Yahia, S., Yu, C.: Interactive itinerary planning. In: 2011 IEEE 27th International Conference on Data Engineering, pp. 15–26. IEEE (2011)
2. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM 2010, pp. 579–588. ACM (2010)
3. Jin, J.S., Xu, C., Xu, M.: A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In: The Era of Interactive Media, pp. 657–670. Springer, New York (2013)
4. Hsieh, H.P., Li, C.T., Lin, S.D.: Exploiting large-scale check-in data to recommend time-sensitive routes. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, UrbComp 2012, pp. 55–62. ACM (2012)
5. Hyoseok, Y., Yu, Z., Xing, X., Woontack, W.: Social itinerary recommendation from user-generated digital trails. Personal and Ubiquitous Computing 16, 1–16 (2012)
6. Ling-Yin, W., Yu, Z., Wen-Chih, P.: Constructing popular routes from uncertain trajectories, pp. 195–203 (2012)
7. Zheng, Y., Zhang, L., Xie, X., Ying Ma, W.: Mining interesting locations and travel sequences from GPS trajectories. World Wide Web Conference Series, pp. 791–800 (2009)
8. Kullback, S., Leibler, R.A.: On Information and Sufficiency. The Annals of Mathematical Statistics 22, 79–86 (1951)

Steganalysis of LSB Matching Based on Image Noise

Jicang Lu, Fenlin Liu, Xiangyang Luo, and Xiaofeng Song

Zhengzhou Information Science and Technology Institute
The State Key Laboratory of Mathematical Engineering and Advanced Computing
450001 Henan Zhengzhou, China
lujicang@sina.com

Abstract. An efficient steganalytic method to LSB matching was proposed in this paper. First, a quantitative measurement method of sensitivity was proposed, and the noise of the cover image is verified to be more sensitive than image content to steganography. Then, extract the higher-order Markov features from the noise of detected images, and distinguish the images with classification tool of Support Vector Machines. Experimental results show that the proposed method outperforms existing methods. Even for low embedding ratio of 0.05 bit per pixel, the detection accuracy can reach up to 79%.

Keywords: steganalysis, steganography, LSB matching, noise, Markov feature.

1 Introduction

Nowadays, the technologies of steganography and steganalysis are hot topics around the world [1]. The covers used by steganography can be various types with redundant data, such as image, audio, and video. For steganography in spatial domain of images, Least Significant Bit (LSB) substitution and $\pm K$ [2] are two typical algorithms. The former is one of the earliest steganalytic algorithms, and is widely used for its imperceptibility and simplicity. However, it can be easily detected with only lower order features [3,4]. Some quantitative algorithms were also presented [5,6,7]. Therefore, it was improved by randomly plus or minus K [8,9] when LSB of cover data should be changed. When $K = 1$, the algorithm is called LSB matching.

On detection of LSB matching, there have been some steganalytic methods based on statistical features, the typical ones includes: center of mass of histogram characteristic function extracted from wavelet domain presented by Ker [10] and Harmsen et al [11], amplitude of histogram presented by Gao and Li et al [12], features extracted based on singular value decomposition (SVD) presented by Gul and Kurugollu [13], noise residual features of decompressed image presented by Zhang et al [14], Markov features of subtractive pixel adjacency matrix (SPAM) presented by Pevný and Fridrich et al [15], and etc. Among these features, the one presented by Pevný and Fridrich et al [15] performs better than others. However, performances of these methods are usually poor when detecting LSB matching, especially for lower embedding ratio. At the some time, almost all of these methods did not take into account influences caused by image contents, which may be one reason of low detection accuracy. Therefore, it is very necessary to design a more efficient detection method.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_31, © Springer-Verlag Berlin Heidelberg 2014

In this paper, a more efficient steganalytic method is proposed based on higher-order Markov features from image noise. Experimental results show that, the proposed method outperforms existing methods, especially in lower embedding ratio.

The rest of the paper is organized as follows. Section 2 will present the sensitivity evaluation method and steganalytic algorithm. In section 3, experiments are reported to verify efficiency of the proposed method. The paper is concluded in Section 4.

2 Steganalysis Based on Image Noise

The quantitative analysis for sensitively of the original image and the noise is first discussed. Then, the steganalytic method for LSB matching is presented.

2.1 Sensitivity Analysis from Qualitative to Quantitative

A natural image can generally be separated into two parts: noiseless image and image noise. Denoting the noiseless image and noise matrix of the natural image as X and N, respectively, the composition of the actual image I can be expressed as:

$$I = X + N \tag{1}$$

The message hiding can be considered as an additive noise process. Denote the embedded additive noise as Ns, then the image after embedding can be expressed as:

$$I' = X + N + Ns \tag{2}$$

The noise (referred to as N') consists of the original noise and the embedded noise: $N' = N + Ns$. For natural images, the noise is generally independent with the image content, and handles the distribution of Gaussian Model. For that noise, the linear filters such as Mean filter are more appropriate to be used for denoising. In this paper, a Mean filter with size 3×3 is used to obtain the noise. In Ref.[15], the Markov features are extracted from the subtractive pixel adjacency matrix. However, the pixels are closely associated with image content. The features of difference matrixes between images with rich content and flat content are rarely similar with each other.

At present, the quantitative analytic method for the sensitivity by is few. Based on the altered histogram of the difference matrix, a quantitative measurement method of the sensitivity is proposed in this paper. Denote difference histograms before and after embedding as $h_c = \{h_c(d) \mid d = ...,-1,0,1,...\}$ and $h_s = \{h_s(d) \mid d = ...,-1,0,1,...\}$, respectively, where $h_c(d)$ and $h_s(d)$ denote the occurrences with difference d of cover and stego image, respectively. Then, calculate mean rate of the altered histogram:

$$MR_h = \frac{1}{2T_d} \sum_{d=-T_d}^{T_d} \left| \frac{h_s(d) - h_c(d)}{h_c(d)} \right| \tag{3}$$

where T_d is threshold for the differences, the results is considered as the sensitivity.

2.2 Sensitivity Comparison between Image Noise and Content

Take differences of right horizontal direction as example, the sensitivities calculated from the image directly and image noise are denoted as I_MR_h and N_MR_h, respectively. The images are from BOWS2lib with 10000 grayscale images and BOSSlib with 9074 grayscale images. The stego-images are constructed with embedding ratio 0.4 bpp (bit per pixel). The threshold T_d is tested on several values from 4 to 8. Then, the sorted differences $N\_MR_h - I\_MR_h$ for each database are shown in Fig.1, and the ratios of images for $N\_MR_h - I\_MR_h > 0$ are listed in Table 1.

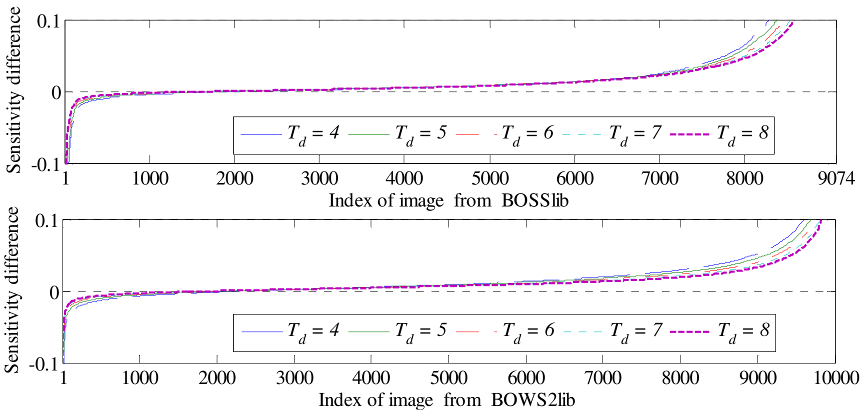


Fig. 1. The differences of the mean altered rates of the difference matrix histogram between detected image and image noise in BOWS2lib and BOSSlib, respectively

Table 1. The ratios of the images that difference matrix of noise matrixes are more sensitive to embedding than that of detected image

Database	$T_d = 4$	$T_d = 5$	$T_d = 6$	$T_d = 7$	$T_d = 8$
BOWS2lib	0.7703	0.7870	0.8006	0.8014	0.8043
BOSSlib	0.7575	0.7739	0.7844	0.7942	0.7970

It can be seen from the experimental results in Fig.1 and Table 1 that, the difference matrix of image noise is more sensitive to embedding than that of the detected image for most of the images.

2.3 Steganalysis Based on Markov Features in Image Noise

Different from [15], this paper will extract features from the noise obtained using 3×3 Mean filter. Similarly, take right horizontal direction "→" as an example (eight directions {→, ←, ↓, ↑, ↖, ↘, ↗, ↙} in all), and denote the difference matrix with size $H\times(W-1)$ of the noises as $D^{\rightarrow}$. Then, second order Markov feature $M^{\rightarrow}$ is:

$$M^{\rightarrow}_{u,v,w} = \Pr\left(D^{\rightarrow}_{i,j+2} = u \mid D^{\rightarrow}_{i,j+1} = v, D^{\rightarrow}_{i,j} = w\right),\ \ u,v,w \in \{-T,..,T\} \tag{4}$$

where $M^{\rightarrow}_{u,v,w} = \Pr\left(D^{\rightarrow}_{i,j+2} = u \mid D^{\rightarrow}_{i,j+1} = v, D^{\rightarrow}_{i,j} = w\right) = 0$ if $\Pr\left(D^{\rightarrow}_{i,j+1} = v, D^{\rightarrow}_{i,j} = w\right) = 0$, and T is the threshold for the features. The final features of eight directions are:

$$F_{1,\,2,\ldots,\,k} = (M^{\rightarrow} + M^{\leftarrow} + M^{\downarrow} + M^{\uparrow})/4 \tag{5}$$

$$F_{k+1,\,k+2,\ldots,\,2k} = (M^{\nwarrow} + M^{\searrow} + M^{\nearrow} + M^{\swarrow})/4 \tag{6}$$

The dimension of the feature is $2k = 2(2\times T + 1)^3$, which is 686 for T = 3. In this paper, the third and forth order Markov features are extracted to capture tiny changes of the noise. Absolute values of the differences will be considered to get 512 and 2048 dimensional features respectively. But it is too high for forth order, let T = 2 to get 486 dimensional features. The features are calculated as follows:

$$M^{\rightarrow}_{u,v,w,x} = \Pr\left(D^{\rightarrow}_{i,j+3} = u \mid D^{\rightarrow}_{i,j+2} = v, D^{\rightarrow}_{i,j+1} = w, D^{\rightarrow}_{i,j} = x\right) \tag{7}$$

where $M^{\rightarrow}_{u,v,w,x} = 0$ if $\Pr\left(D^{\rightarrow}_{i,j+2} = v, D^{\rightarrow}_{i,j+1} = w, D^{\rightarrow}_{i,j} = x\right) = 0$, and $u,v,w,x \in \{0,..,T\}$.

$$M^{\rightarrow}_{u,v,w,x,y} = \Pr\left(D^{\rightarrow}_{i,j+4} = u \mid D^{\rightarrow}_{i,j+3} = v, D^{\rightarrow}_{i,j+2} = w, D^{\rightarrow}_{i,j+1} = x, D^{\rightarrow}_{i,j} = y\right) \tag{8}$$

where $M^{\rightarrow}_{u,v,w,x,y} = 0$ if $\Pr\left(D^{\rightarrow}_{i,j+3} = v, D^{\rightarrow}_{i,j+2} = w, D^{\rightarrow}_{i,j+1} = x, D^{\rightarrow}_{i,j} = y\right) = 0$, $y \in \{0,..,T\}$.

Efficiency of the proposed feature will be experimentally verified in next section.

3 Experiments and Results

The experiments are carried out based on the following two image databases:

(1) BOWS2lib: this database was used during BOWS2 contest[1], and contains approximately 10, 000 grayscale images with fixed size of 512×512 coming from rescaled and cropped natural images of various sizes.

(2) BOSSlib: this database was recently constructed for the first challenging on steganalysis[2], and contains approximately 9, 074 grayscale images with fixed size of 512×512 coming from rescaled and cropped raw images of various sizes.

Construct stego-images using LSB matching steganography with embedding ratios include 0.05, 0.1, 0.2 and 0.4 bpp. To sum up, there are 95, 370 images. For the classifier training, randomly select 5,000 of each group in BOWS2 and 6,000 of each group in BOSS as image database. The remaining images are used for performance testing. SVM is used to train the classifier.

---

[1] `http://bows2.ec-lille.fr/`

[2] `http://www.agents.cz/boss/BOSSFinal/`

In this paper, the 2nd, 3rd and 4th Markov features were all tested. Their thresholds T were set to 3, 3 and 2, respectively. The evaluation results (mHOM denotes the proposed method) are listed is Table 2.

It can be seen from Table 2 that, for embedding ratio 0.05, 0.1 and 0.2 bpp, the detection accuracy of the proposed method is superior to SPAM. Even for embedding ratio of 0.05 bpp, the detection accuracy can reach up to 79%. And for ratio 0.4 bpp, the result is comparative with SPAM.

Table 2. Comparison of the classification results between SPAM and the method proposed

Embedding ratio	Order of the feature	BOWS2lib		BOSSlib	
		SPAM	mHOM (Proposed)	SPAM	mHOM (Proposed)
0.05 bpp	2nd	76.41%	74.26%	77.72%	77.62%
	3rd	73.69%	76.82%	74.72%	78.38%
	4th	74.84%	**78.43%**	77.10%	**79.00%**
0.1 bpp	2nd	83.35%	82.41%	83.69%	84.92%
	3rd	81.43%	84.05%	81.18%	84.61%
	4th	81.41%	**84.88%**	82.32%	**84.99%**
0.2 bpp	2nd	89.84%	88.81%	89.38%	**90.53%**
	3rd	88.38%	90.24%	87.36%	89.02%
	4th	88.31%	**90.69%**	86.87%	89.46%
0.4 bpp	2nd	**94.50%**	94.21%	**93.82%**	93.66%
	3rd	93.00%	93.89%	90.83%	92.89%
	4th	92.78%	94.28%	90.09%	92.36%

At the same time, evaluate the algorithm for combination of the image databases, which includes inter-database combination and intra-database combination. The former is to combine images of the same database, such as BOWS2lib or BOSSlib. Inter-database combination is to combine all the images from BOWS2lib and BOSSlib with all embedding ratios. The experimental results can be obtained as Table 3.

Table 3. Comparison of the classification results between SPAM and the method proposed when detecting the combination of different image databases with various embedding ratios

Database	SPAM_2nd	mHOM_4th
BOWS2lib	82.83%	**84.18%**
BOSSlib	83.81%	**85.07%**
Combining	83.98%	**85.27%**

The results in Table 3 reflect that, the proposed algorithm mHOM with 4th order Markov features performs better than SPAM with 2nd order Markov features fro both inter- and intra-database combination.

The results above show that, the performance of the proposed method outperforms existing method, especially for lower embedding ratio, which indicate that the steganalytic method proposed in this paper is more effective in practical detection.

4 Conclusions

In this paper, a more efficient detection method of LSB matching is proposed based on higher-order Markov features extracted from image noise. The image noise is obtained by using the Mean filter which performs well when denoising the Gaussian noise. Experimental results show that the proposed method with Mean filter outperforms previous typical methods.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (61272489, 61379151 and 61302159), and the Doctoral Dissertation Innovation Fund of Zhengzhou Information Science and Technology Institute (BSLWCX201203).

References

1. Ker, A.D., Bas, P., Böhme, R., Cogranne, R., Craver, S., Filler, T., Fridrich, J., Pevný, T.: Moving steganography and steganalysis from the laboratory into the real world. In: ACM Workshop on Information Hiding and Multimedia Security, pp. 45–58. ACM Press (2013)
2. Alturki, F., Mersereau, R.: A novel approach for increasing security and data embedding capacity in images for data hiding applications. In: International Conference on Information Technology: Coding and Computing, pp. 228–233. IEEE Press (2001)
3. Westfeld, A., Pfitzmann, A.: Attacks on steganographic systems. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 61–76. Springer, Heidelberg (2000)
4. Yang, C.F., Liu, F.L., Luo, X.Y., Zeng, Y.: Pixel group trace model-based quantitative steganalysis of multiple least significant bits Steganography. IEEE Transactions on Information Forensics and Security 8(1), 216–228 (2013)
5. Fridrich, J., Goljan, M., Rui, D.: Detecting LSB steganography in color and grayscale images. IEEE Transactions on Multimedia 8(4), 22–28 (2001)
6. Dumitrescu, S., Wu, X.L., Wang, Z.: Detection of LSB steganography via sample pair analysis. IEEE Transactions on Signal Processing 51(7), 1995–2007 (2003)
7. Luo, X.Y., Liu, F.L., Yang, C.F., Lian, S.G.: Modification ratio estimation for a category of adaptive steganography. Science China: Information Sciences 53(12), 2472–2484 (2010)
8. Li, X.L., Yang, B., Cheng, D.F., Zeng, T.Y.: A generalization of LSB matching. IEEE Signal Processing Letters 16(2), 69–72 (2009)
9. Luo, W.Q., Huang, F.J., Huang, J.W.: Edge adaptive image steganography based on LSB matching revisited. IEEE Transactions on Information Forensics and Security 5(2), 201–214 (2010)
10. Ker, A.D.: Steganalysis of LSB matching in grayscale images. IEEE Signal Processing Letters 12(6), 441–444 (2005)

11. Harmsen, J.J., Pearlman, W.A.: Steganalysis of additive noise modelable information hiding. In: The SPIE, Security, Steganography, and Watermarking of Multimedia Contents V. SPIE, vol. 5020, pp. 131–142. SPIE (2003)
12. Gao, Y.K., Li, X.L., Yang, B., Lu, Y.F.: Detecting LSB matching by characterizing the amplitude of histogram. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1505–1508. IEEE Press (2009)
13. Gul, G., Kurugollu, F.: SVD-based universal spatial domain image steganalysis. IEEE Transactions on Information Forensics and Security 5(2), 349–353 (2010)
14. Zhang, J., Zhang, D.: Detection of LSB matching steganography in decompressed images. IEEE Transactions on Information Forensics and Security 5(2), 141–144 (2010)
15. Pevný, T., Bas, P., Fridrich, J.: Steganalysis by subtractive pixel adjacency matrix. IEEE Transactions on Information Forensics and Security 5(2), 215–224 (2010)

Development of Automotive Multimedia System Using Visible Light Communications

Sang-Yub Lee*, Jae-Kyu Lee, Duck-Keun Park, and Sang-Hyun Park

Ballyong-ro, Deokjin-gu, Jeonju-si, Jeolabuk-do, 561-844
Korea Electronics Technology Institute, Republic of Korea
{syublee,jae4850,parkdk,shpark}@keti.re.kr

Abstract. Among the variety of vehicle technology trend issues, the biggest one is focus on automotive network system. Especially, optical network system is preferred. The automotive application considered as a service which can be provided to users inside a car such as a media playing or data transmission. For the media playing service needed to transfer high data rate and high display quality, it is required least scores of Mbps to be formed communication link. In this work, it is introduced that VLC (Visible Light Communications) system interlocked with MOST (Media Oriented System Transport) which is commercialized optical network system for transmission of automotive multimedia data has been implemented the network interconnection. The main advantage of proposed system is the robustness against EM (Electromagnetic) interference and interruption of streaming data in car.

Keywords: In-vehicle network systems, MOST, VLC, Multimedia system.

1 Introduction

With the passage of time, vehicle technology in modern times has been developed rapidly. Recent of today's automobile research area has been gradually changed from mechanics to electronics to the way of offering entertainment service to customs that can serve the connection to smart device easily and conveniently.

In particular, most of people want to be experienced in high quality multimedia data service while in driving. According to customer's demands, MOST system is developed to provide proliferate and cost effective fabric to transmit data between some devices attached to the harsh environment of automobile. Lately, conventional car media system which has been adopted to wireless communications system such as Bluetooth or Wireless LAN in order to transit the media data source, but automotive streaming environment is tough to exchange streaming data without ceasing link connection. Proposed VLC system does not interfere with radio frequency systems and avoiding EM compatibility restrictions, thus it can be the only solution for streaming service tool in car. Additionally, as the distinctiveness of light, designed system is secure against intentional attempts of unauthorized access. Forced to network

* Jeonbuk Embedded System Research Centre.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_32, © Springer-Verlag Berlin Heidelberg 2014

connectivity, in this paper, we tried that it linked with automotive optical network MOST. Thus, the network management of MOST is used to same mechanism in this system. In Section 2, MOST-VLC network system and their data processing are introduced for vehicle environment. The VLC interface module which has been implemented with MOST networks that propagates the media data in air described in Section 3. In Section 4, platform and demonstration of streaming service are shown. This paper is concluded in Section 5.

2 MOST-VLC Network System

MOST is the de-factor standard for efficient and cost effective networking of automotive multimedia and infotainment system [1]-[2]. The current MOST standards released MOST150, 150 means that 150Mbps network bandwidth with quality of services is available. To meet the demands from various automotive applications, MOST network system provides three different message channels: control, synchronous used in streaming service and asynchronous channel only for packet data transfer. In describe in Fig.1, proposed network system is consist of MOST devices with VLC module as a ring network. As it has been discussed before [3], conventional VLC system had limitations of data bandwidth. It was a maximum 2 Mbps Visible light downlink and 125Kbps infrared uplinks such as light sensors. But, proposed MOST-VLC network system which is linked with MOST network system has the same as those data rate. It can be 150Mbps uplink and downlink that make a variety of utilizations as a means of internal wireless communications in car.

Fig. 1. MOST-VLC Network System

2.1 MOST-VLC Data Processing

For the data processing, it is explained in streaming data part and network service one. Being presented the streaming data part; the bandwidth of the streaming data channel can be calculated using the following formula:

$$\text{Bandwidth} = 372\text{bytes} \times 8\text{bit} \times 48\text{kHz} \tag{1}$$

Synchronous data is used for the real-time transmission of data. Before data transmission, the data transmitted for the synchronous connection must be established by the connection master in network system. For this method, one socket of VLC connected with MOST has to be created and connected at the interfaces to the frame and to the local resource depicted in Fig. 2. The content of the frame remains unmodified until the frame arrives back at the sending node. The quasi-static establishment of connections on a channel is denominated as TDM (Time Division Multiplexing). The data are transmitted cyclically in a specified time pattern at the same frame position. There is no repetition in the case of communication errors. A valid value is then available the next cycle.

In addition to aspect of network service, the net-block that is responsible for the administration of a device must be implemented in VLC device. Thus, VLC unit is usually found together in the head unit. Net-block has a list of all the function blocks implemented on the device, and manages all their network address in ring network.

Fig. 2. The Data Process of MOST-VLC system

3 VLC Interface Module

3.1 Processing Module for VLC Interface

Within a VLC device, the processing module possesses the role of a two way bridge between opto-electronic transceiver and the interior network of the device. Fig. 3 shows an example for the implementation of a typical processor style VLC device. The VLC device provides the transmission of converting data between optical signal and electrical one. Simultaneously, this allows sending and receiving the MOST data via EHC (External Host Controller) in an efficient way.

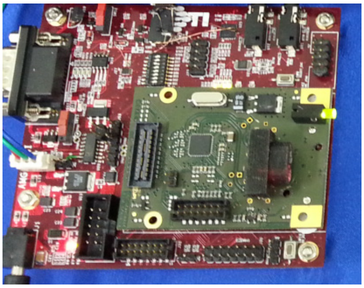


Fig. 3. VLC interface module

3.2 Physical Layer of VLC Optic Interface

The physical layer of VLC module is characterized as in Fig.4. It represents the physical connection between two VLC control components. Similar to MOST physical layer, four specification points SP1, SP2, SP3 and SP4 are defined. The SP1 and SP4 describe the electrical signal requirements between a VLC interface and a converter.

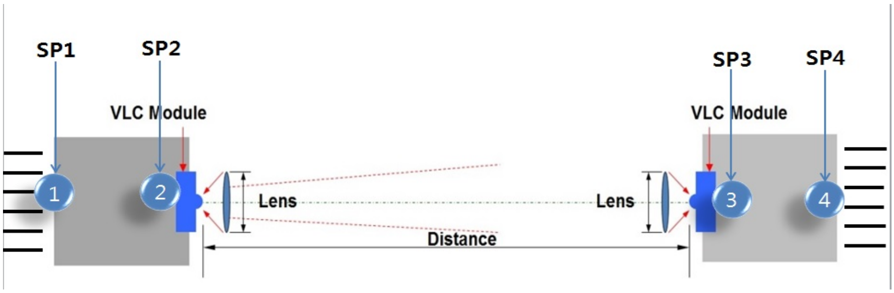


Fig. 4. VLC optic interface

And the wireless interface properties between VLC devices are showed by SP2 and SP3. For SP2 and SP3, opto-electronic transceivers LED diodes having an emission wavelength of 650nm are used. SP1 and SP4 including driver component and amplifier as the case may be, are attached onto leadframe and encapsulated in a transparent plastic material. Six connecting pins lead out of the plastic housing to provide the contacts for current supply, differential data signals and other status or control lines. Especially, to execute the wireless communication with VLC module, beam focusing lens is developed in Fig. 5.

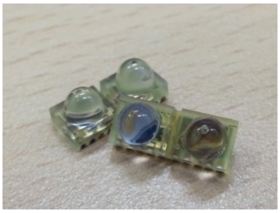


Fig. 5. Beam focusing lens

4 Implementation of Streaming Service

4.1 System Architecture

The basic system architecture of MOST-VLC unit, as it shown in Fig. 6, can be deduced from Fig. 2. It shows a device with an MOST optic interface as MOST side. The MOST optic interface transforms the light signal into an electrical signal which transfers the data routing module. When processing data routing, MOST frame involved in control, streaming and packet transmitted to Head Unit via MLB. And, the data processing is realized in accessing to the VLC interface module. Head Unit makes data processing control master clock cycles for synchronization. According to the required timing, VLC optic interface forms the optical signal and sends media data in wireless channel.

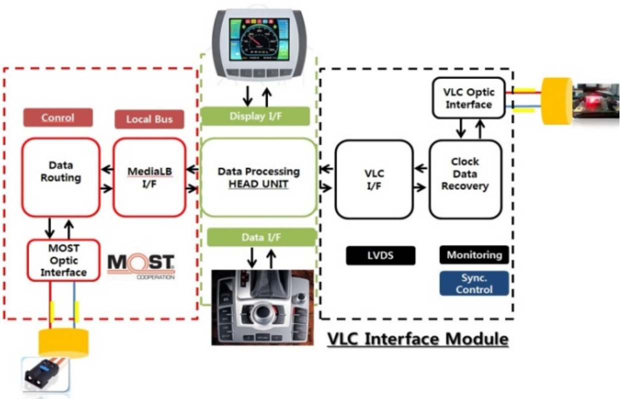


Fig. 6. Proposed system architecture of MOST-VLC

4.2 System Synchronization

An important part of the proposed system for signal handling on the communication side between VLC and MOST networks is synchronization. Each of interface controller either creates the system clock (when being configured as the sole master of the system), or synchronizes to the incoming bit stream. When the master device receives the frame again when it has traveled around the ring, it reclaims the signals of a PLL connection and subsequently generates the next frame.

4.3 Streaming System Demonstration

As The MOST-VLC application framework permits described in Fig. 7, the creation of applications that are independent of a specific implementation on a certain device and those of the other applications with which they interact in a MOST network system. This paper describes the aspects of the audio streaming service that VLC device sends audio sources stored in user's handheld unit to VLC receiving device interlocked with Head Unit that operates MOST networks management. The protocol that audio applications use to exchange data in a MOST system is described next and streaming system demonstration is represented in Fig. 8.

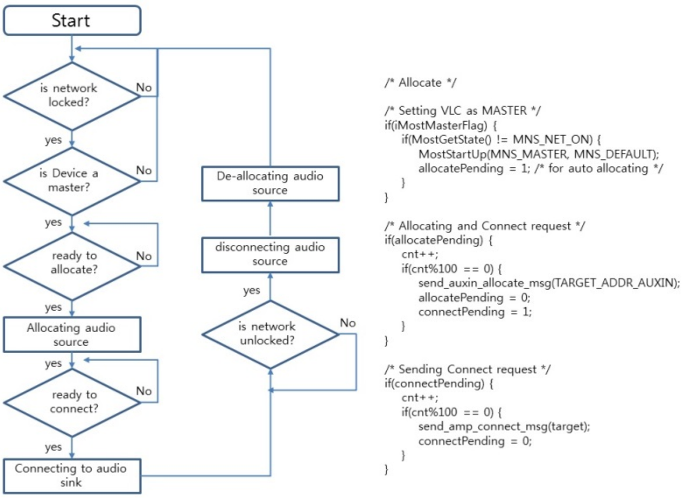


Fig. 7. Application Framework of VLC device

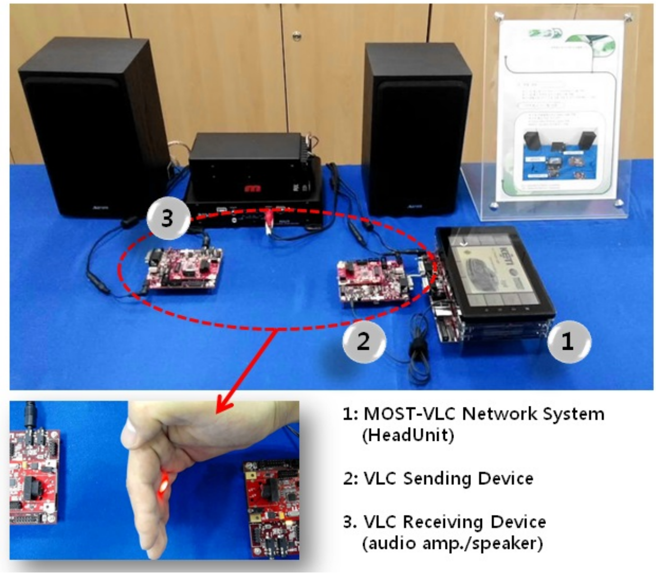


Fig. 8. Streaming System Demonstration

5 Conclusion

For the trend of car infotainment system is moving to the high quality audio sound system, this paper is introduced the development of the audio streaming service based on wireless optical network system called as VLC. Particularly, MOST-VLC, to be optimized streaming data transmitting without ceasing communication on vehicle environment is satisfied with reducing the weight and ensuring the reliability for the free of electro-magnetic problems. With optical wireless device, designed network platform realized data transmission over 100Mbps class and demonstrated that self-contained VLC speaker and woofer can be served to passengers more conveniently. Especially, outstanding point for implementation of wireless audio streaming service is represented to hardware architecture and software frame depending on network connectivity with MOST networks.

References

1. Grzemba, A.: MOST Books from MOST25 to MOST150. MOST Cooperation. Franzis (2012)
2. Strobel, O., Rejeb, R., Lubkoo, J.: Communication in Automotive System Principles, Limits and New Trends for Vehicles, Airplanes and Vesseles. In: IEEE ICTON, pp. 1–6 (2007)
3. Rufo, J., Quintana, C., Delgado, F., Rabadan, J., Perez-Jimenez, R.: Considerations on modulations and protocols suitable for visible light communications channels. In: IEEE CCNC, pp. 362–364 (2011)

Research and Implementation of CAD-INP Integration System Based on OpenCASCADE

Qingguo Zhou[1,*], Shuan He[1], Jiaming Yang[1], Hubin Yang[1], Yanshan Tian[1,3], Junqiong Wang[1], Lei Yang[2], and Ji Qi[2]

[1] School of Information Science & Engineering, Lanzhou University, Lanzhou, Gansu, China
{zhouqg,hesh12,yangjm12,tianysh12,wangjunqiong}@lzu.edu.cn,
Yhbczj999@gmail.com

[2] Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, Gansu, China
{lyang,qijiin21c}@impcas.ac.cn

[3] School of Mathematics & Computer Science, Ningxia Teachers University, Guyuan, Ningxia, China
tianysh12@lzu.edu.cn

Abstract. MCNP is a general-purpose Monte Carlo N-Particle code that can be used for neutron, photon, electron, or coupled neutron/photon/electron transport. MCNP input file (INP file) has the characteristics of complicated form and is error-prone in describing geometry model. This paper proposes a method combined both geometric modeling and converting the models to INP file, the method realized a system to integrate three-dimensional modeling and converting. The three-dimensional modeling part developed based on open geometry engine OpenCASCADE, and it can export STEP(Standard for the Exchange of Product Model Data) file directly. The converting part accomplished the converting from STEP file to INP file.

Keywords: MCNP, 3D modeling software, OpenCASCADE, STEP, INP.

1 Introduction

By performing the nuclear analyses with the Monte Carlo code MCNP and the detailed calculation geometry data, nuclear properties such as neutron fluxes, nuclear heating, tritium production rate dose rate, can be obtained with high accuracy in nuclear design of nuclear fusion reactors [1]. So the study of geometrically processing function on MCNP is constantly a research focus. For input file of MCNP, named the INP file [2], describes the complicated geometrical model data, has always been an important part in MCNP research. For an identical physical model, even though the scale or complexity of the model is extremely small, the description method between INP files and other three-dimensional modeling software has a great difference. If using the existing commercial three-dimensional modeling software to construct a geometric model, the transformation from geometric model into INP description

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_33, © Springer-Verlag Berlin Heidelberg 2014

model will be an indispensable and complicated procedure. Up to now, there are many researchers who have conducted a lot of studies on the transformation from the model constructed by three-dimensional modeling software to MCNP geometric model. For instance, in [3], Zhang jiansheng proposes a transformation algorithm for transforming UG model to MCNP geometric model. And in [4], Zhang makes the improvement of the algorithm.

Usually, commercial CAD systems have their own APIs. Unigraphics has OpenAPI, ProEngineer has Pro/Toolkit, OpenDIS is the API of DINA-CAD [5]. With these APIs， we can conduct a secondary development based on commercial modeling software, which is helpful for modeling and converting task. After a period of time on development based on such modeling software, we find that the function designing, data transferring and data exchanging are limited because commercial three-dimension modeling software always only takes its customers into consideration. This is one of the reasons why we propose to develop our own three-dimension software in this paper. Another reason is that using three-dimension software developed by ourselves, we can clearly understand the construction, description and process of creating geometric models. Understanding the details of geometric model constructing procedure will be greatly helpful for the following work, transforming common model to INP geometric model.

OpenCASCADE is released by Matra Datavision Company of France in 2000 [6]. It integrates wire frame modeling, surface modeling and solid modeling and provides an open geometric modeling platform for all kinds of 2D/3D modeling application development [7]. Furthermore, OpenCASCADE also provides a wide set of topological algorithms which allowing to create and modify topological objects using different construction techniques [8]. Based on these advantages, FreeCAD, NaroCAD, HekksCAD, Total Engineer and other analogous three-dimension modeling software are developed. In this paper, we design a 3D modeling software based on the OpenCASCADE kernel. It is developed especially for MCNP geometric model.

STEP, the Standard for the Exchange of Product Model Data, is an international product data standard to provide a complete, unambiguous, computer-interpretable definition of the physical and functional characteristics of a product throughout its life cycle. It is a much broader standard than data interchange standards such as IGES. STEP is already widely distributed in the domain of mechanical engineering and gets more and more support also in the aerospace domain [9]. As some kinds of neutral file, STEP is accepted and supported by the main commercial three-dimension modeling software in this field [10-14]. So lots of researches have been done on the transformation from STEP files into INP files. These useful studies have achieved the transformation from geometric model which is described by STEP files into INP geometric model.

2 The Architecture of CAD-INP Integration System

The detailed system structure is shown in figure 1.The big arrow in figure 1 represents the data flow direction. It can be seen from figure 1 that the user and 3D modeling software can interact with each other. It is because the 3D modeling environment provides the interactive interface of the input data for the users to achieve the aim of the interaction. The data stream between the 3D software modeling and the INP input document is unidirectional because of the single direction of data transformation which is from the STEP files of the 3D modeling software output to the INP input files. Since MCNP can recognize the INP input files, the INP files obtained from the previous work can be used in the MCNP directly.

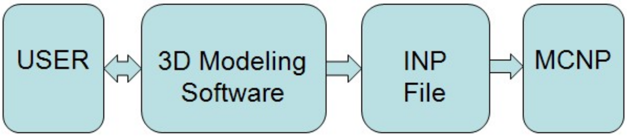


Fig. 1. Architecture of CAD-INP Integration System

2.1 3D Modeling Software

As the biggest section of the system proposed in this paper, the construction of the 3D modeling software is complicated. The software architecture is shown in figure 2.

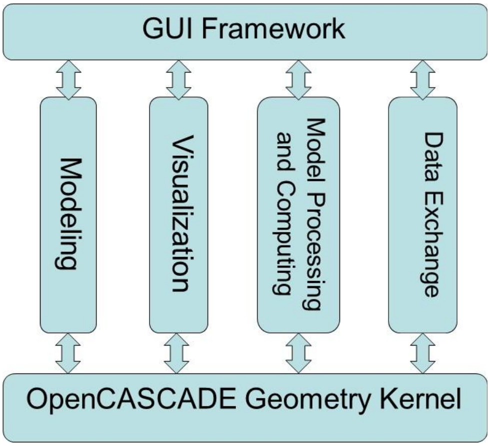


Fig. 2. Architecture of Modeling Software

It can be seen from the figure 2 that 3D software is divided into three layers. The OpenCASCADE Geometry Kernel, a powerful 3D modeling libraries, is in the most-derived. It provides the basis API of the point, line, face, body as well as the complex form display and interaction which can complete basic geometric model creation. After the depth development, OpenCASCADE Geometry Kernel can realize the texture, illumination, primitive filling and rendering graphics operations as well as amplification, narrow, rotation, roaming, flight simulation, penetration simulation dynamic effect and so on.

Then, the middle layer which including four parts is the core of modeling software. The first part is the modeling part which implements the fundamental point, line, face, and the creation of the basic voxel. The second one is the visualization part which is the most basic module. When the geometric model is created, it needs to be displayed. The process of the model operation should be preceded in a visual environment. The third part is model processing and computing which implements the intersection, as well as summation operation among the model entities. This part provides the gap, overlap repair operation after model established. The last part is the data exchange part. This software can exchange data with other CAD software by using the data exchange interface which ensures the data sharing. At present, the data exchange standards include STEP, IGES, STL, VRML, and so on.

The uppermost layer is GUI framework of the CAD modeling software which can be used to interact with users directly. This part integrates the visible functions which can display the user interactive operations, model visualization on the CAD software interface. Meanwhile, this layer also completes the data interaction between the user and the software.

2.2 From 3D Modeling Software to INP Input File

The 3D modeling software implements the data exchange part, so the created geometrical model can be saved in the STEP file outside of the program. This implement can reserve the geometry entity model from the computer memory to the external disk. The STEP file obtained from the software proposed in this paper can be identified and opened by other commercial three-dimensional modeling software because of the general international standards. Since it has this universality, the second part in this system, that is the data exchange part, puts this STEP file as input data. The INP file can be obtained with the specific algorithms which are aiming at information extraction and analysis of the input STEP file.

3 The Implementation of CAD-INP Integration System

The system proposed in this paper is developed on the Linux operating system. The construction of the development platform not only needs the OpenCASCADE geometry engine library, but also needs some other 3rd-party products, such as Freeimage, Freetype, Ftgl, gl2ps, Qt, TBB, Tcl/Tk. 3D modeling software which can be developed after the construction of the development platform, comprises creation of the geometric models, switch display modes of the model, importing and exporting of the models and so on. In order to achieve the goal of implementing the data interaction between software and the user, all the geometrical model parameters can be inputted by users in the modeling creation part. There is an example shown in figure 3.

As we can see from figure 3, it is the parameter input dialog which used to establish the basic voxel ball. In general, four parameters are mandatory in order to establish a ball, the 3D coordinate x, y, z and the radius r of the ball. In the way implemented in this system, users can create the geometrical model of the ball by inputting the four parameters. One complicated model built by the software was shown in figure 4. The whole model can be reserved in the STEP files through the data exchange after the establishment of the model. Then, the INP input file can be obtained from the processing on the STEP file. The descriptive approach of a ball in INP file is shown in figure 5. The radius and the coordinate of the sphere center of the ball is 50.0 and (0.0, 0.0, 0.0), respectively.

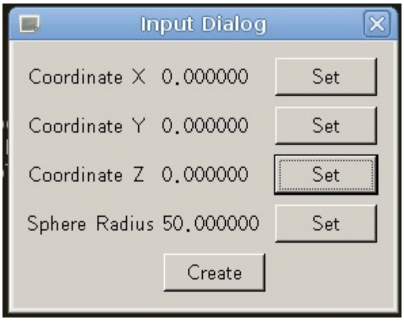


Fig. 3. Parameters Input Dialog

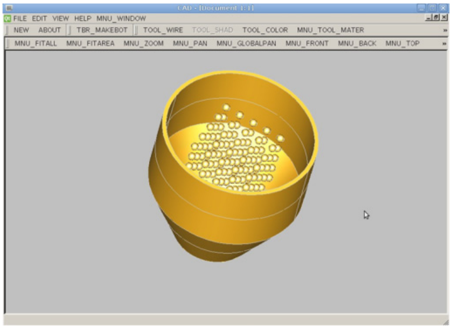


Fig. 4. Complicated Model

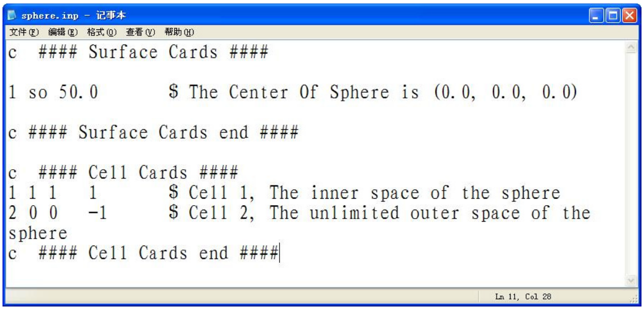


Fig. 5. Content of shpere.inp

4 Conclusion and Future Work

This paper introduces a CAD-INP integration system. Building a platform for developing the modeling software firstly, and then creating models based on OpenCASCADE open source geometry engine library, finally converting the results to the MCNP geometry model. This work constructs the whole creation process of geometric models for MCNP input file, which will promote the efficiency and accuracy of MCNP application. Meanwhile there are more efforts can be made, such as further developments on OpenCASCADE kernel, simulating on traveling through. As for the MCNP geometric model conversion section, the next task is achieving the complex geometric structures and repeated structures model transformation.

Acknowledgements. This work is supported in part by the “Strategic Priority Research Program” of the Chinese Academy of Sciences with Grant No. XDA01020304 and the China National Natural Science Foundation with Grant No. 91026005, National Natural Science Foundation of China under Grant No.60973137, Program for New Century Excellent Talents in University under Grant No.NCET-12-0250, Gansu Sci.&Tech. Program under Grant No.1104GKCA049 and 1204GKCA061, Google Research Awards and Google Faculty Award.

References

1. Sato, S., Nashif, H., Masuda, F.: Progress of conversion system from CAD data to MCNP geometry data in Japan. J. Fusion Engineering and Design 85(7), 1546–1550 (2010)
2. Los Alamos National Laboratory, `http://mcnp.lanl.gov`
3. Huachuan, Z., Cai, J., Huang, Y.: Research and Implementation of an Algorithm of Converting UG Model to MCNP Geometry Model. J. Computer Applications and Software 8, 007 (2009)
4. Zhang, J., Chai, Y.: Studying and Implementing the Improved Algorithm of Converting UG Model to MCNP Geometry Model. J. JisuanjiYingyongyuRuanjian 27(11), 65–67 (2010)
5. Kim, J., Han, S.: Encapsulation of geometric functions for ship structural CAD using a STEP database as native storage. J. Computer-Aided Design 35, 1161–1170 (2003)
6. Open CASCADE Technology, `http://www.opencascade.org`
7. Yan, C., Jiang, D., Lina, Y.: Development of a Sheet Metal Part Stock Layout System Based on OPENCASCADE Platform. In: 1st IEEE International Multi-Symposiums on Computer and Computational Sciences, Hangzhou, pp. 437–441 (2006)
8. Jianqing, Y., Jiangang, L., Yunjiang, L.: NC Machine Interference and Collision Checking System Based on OpenCASCADE. In: 2010 IEEE International Conference on ICDMA, Changsha, pp. 976–979 (2010)
9. Langwost, A., Sdunnus, H., Gunia, D.: Presentation of the PC Version of the Esabase/debris Impact Analysis Tool. In: Proceedings of the 4th European Conference on Space Debris, pp. SP–587. ESA, Darmstadt (2005)
10. Yuetong, L.: Research on Models Conversion and its application in MCNP Model. Doctoral Dissertation (2005)
11. Mingzhu, D.: Research of CAD with MCNP format conversion. Master’s thesis (2008)
12. Sivakumar, S., Dhanalakshmi, V.: An approach towards the integration of CAD/CAM/CAI through STEP file using feature extraction for cylindrical parts. J. International Journal of Computer Integrated Manufacturing 26(6), 561–570 (2013)
13. Shapiro, V., Vossler, D.L.: Separation for boundary to CSG conversion. J. ACM Transactions on Graphics (TOG) 12(1), 35–55 (1993)
14. Buchele, S.F., Crawford, R.H.: Three-dimensional halfspace constructive solid geometry tree construction from implicit boundary representations. J. Computer-Aided Design 36(11), 1063–1073 (2004)
15. Teissandier, D., Delos, V.: Algorithm to calculate the Minkowski sums of 3-polytopes based on normal fans. J. Computer-Aided Design 43, 1567–1576 (2011)
16. Riou, A., Mascle, C.: Assisting designer using feature modeling for lifecycle. J. Computer-Aided Design 41, 1034–1049 (2009)
17. Convard, T., Bourdot, P.: History based reactive objects for immersive CAD. In: Proceedings of the Ninth ACM Symposium on Solid Modeling and Applications, pp. 291–296. Eurographics Association, Genoa (2004)

Nomadic Work Life Support Using ICT: Toward Multi-generational Techno-socio Innovation

Toshihiko Yamakami

ACCESS, Software Solution,
1-10-2 Nakase, Mihama-ku, Chiba-shi, Japan 261-0023
http://www.access-company.com/

Abstract. Innovation explores untapped intellectual human resources, as demonstrated during the Industrial Revolution and in the globalization. In the 21st century, the final goal of such innovation is to redevelop the human resources that senior citizens can provide. The author proposes a multi-generational collaborative framework that utilizes these senior resources.

1 Introduction

We have witnessed many breakthroughs in the area of technology through the past decades. Technology has a great impact on innovation when manufacturing becomes a critical bottleneck.

Even with the illuminating success of advances in technology, it is important to capture the essential part of innovation, which lies in the exploration of human resources. Drucker argued that many of the job increases during the middle of the 20th century were attributable to the innovation of management in medium-size enterprises. The industrial revolution created urban factory workers out of what were previously farmers. Globalization is shaking the current landscape, which is an exploration of the work-forces in the developing countries, and is supported by advances in ICT (information and communication technology).

From this viewpoint, the next opportunities for innovation lie in the exploration of senior human resources. The aging trend leverages the number of senior citizens worldwide.

The author examines the multi-generational collaboration from the viewpoint of a new type of innovation.

2 Background

2.1 Purpose of Research

The aim of this research is to identify a multi-generational collaborative framework that enables innovation through the exploration of senior human resources.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_34, © Springer-Verlag Berlin Heidelberg 2014

2.2 Related Works

Chesbrough discussed open innovation for today's trends [1]. Gratton presented two views on the future of work from ongoing technological and demographic changes [2].

The author Yamakami presented the trend of cross-boundary software engineering from the viewpoint of open innovation [3]. He also discussed virtual world alchemy for creating new values from service engineering [4].

Past research addressed ICT support for senior citizens. However, it lacked the viewpoint of innovation that utilizes the unexplored potential of senior citizens through ICT. The originality of this paper lies in its examination of innovation in cross-boundary engineering of multi-generational collaboration.

3 Trends

3.1 Background: Aging

Issues attributable to population aging are depicted in Table 1. It is one of the most critical issues of the 21st century in the developed countries.

Table 1. Issues attributable to population aging

Issue	Description
Cost of pension	Budget deficit from increasing pension costs
Cost of medicare	Budget deficit from increasing medicare costs for senior citizens
Decreased economic growth	Decreased labor population leading to decreased economic growth
Disparity of wealth	Seniors exhibit a greater disparity of wealth than younger Generations
Conservatism	Increasing the number of senior citizens leads to increased political conservatism

This is causing changes in the nature of the nation, with demographic changes, from salary workers to pensioners. This trend indicates that most nations need to address the issues attributable to aging, because they are the inevitable outcome of the improved public health conditions worldwide.

3.2 Techno-socio Innovation

The trend of innovation is depicted in Fig. 1.

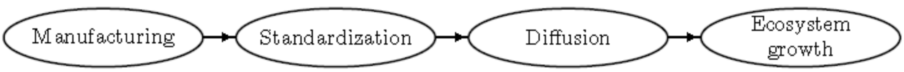


Fig. 1. Trend of mainstream innovation

In the past, most innovation focused on manufacturing, because it was the hardest part of the challenge. With the high-speed sharing of information, that past innovation is no longer applicable in many cases. Standardization played an important role in innovation, in many cases in the development of communication products. Diffusion was a bottleneck, for example, in the case of wireless TV telephony and MMS (the multimedia version of SMS). Ecosystems showed key aspects, for example, in the cases of the AppStore and Google Play. Ecosystem engineering inevitably leads to the engineering of business models and platforms for heterogeneity. It leverages the need for cross-border engineering and open innovation. However, it needs to address usability, learnability, constructability, and distributability also. Platform engineering has to cope with persuasion, influence, and expectation engineering. Many of these features are relatively new to the engineering field. This has triggered shifts from technical innovation to social innovation, from unification to multiplication, and from manufacturing to usage and services.

The parsing of past innovation is depicted in Table 2. From this viewpoint, the only unexplored human resources on the scale of tens to hundreds of millions on this planet are senior citizens. The author has come to recognize this to represent the last opportunity for innovation in the 21st century.

Table 2. Parsing of the past innovation

Example	Description
Industrial revolution	Turning farmers into industrial workers with the technology of mass production.
Globalization	Turning the human resources of developed countries toward utilization of emerging ICT.

3.3 Collaboration

A two-dimensional view model of the enhancement of technology-augmented collaboration is shown in Fig. 2. Multi-generational collaboration is an attempt to take a route in the direction of the dimension of Society.

The trend of collaboration with an increasing diversity is shown in Fig. 3.

The trend indicates a shift toward increased flexibility and heterogeneity. One clue to support such soft collaboration is the support that ICT has for virtual teams.

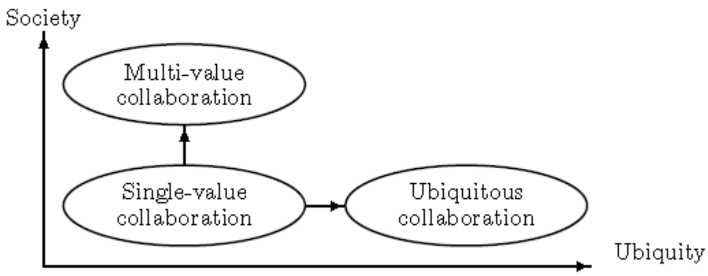


Fig. 2. Two-dimensional view of technology-augmented collaboration

Fig. 3. Trend of collaboration with an increasing diversity

4 Nomadic Work-Life for the Next Generation of Social

The reincarnation of the nomadic life into a modernized nomadic life is depicted in Fig. 4.

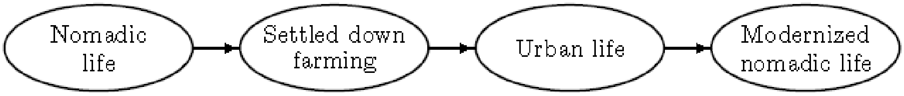


Fig. 4. Reincarnation into a modernized nomadic life

The settled-down urban life fits for highly mobile nuclear families are frequently seen in the trend of high-growth society during industrialization. This does not provide a good fit for an aging society however, because urban cities were not originally built for senior citizens. Even part of the Tokyo metropolitan area was rural from today's standards only a century ago.

A triangular view model of multi-generational collaboration is depicted in Fig. 5. This model puts senior working holidays as the core for a shift in work-life balance. Today's industrial environment does not require much physical strength. Senior pensioners do not need salary income, but they do need everyday work-life targets. This can be accomplished by a nomadic senior work life with multi-generational collaboration supported by ICT.

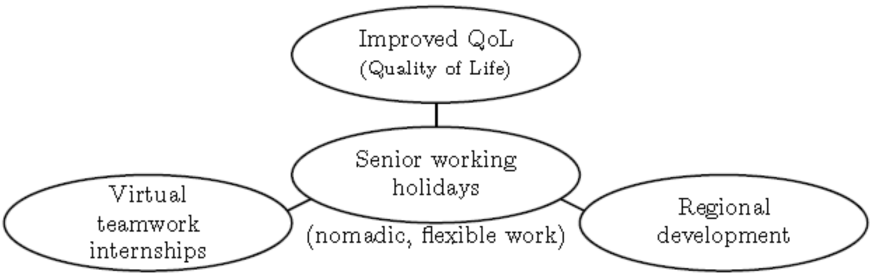


Fig. 5. Triangular view model of multi-generational collaboration with senior working holidays

Support for a nomadic senior work life requires both technical and social solutions. Technical solutions that support a nomadic senior work life are depicted in Table 3.

Table 3. Possible technical measures that support nomadic senior work life

Function	Description
Remote medicare	Share medicare records are shared anytime and anyplace to reduce concerns about remote work-life.
Internet green card	People hate stateless wanderers. Stay-records and job records provide the social proof.
Virtual team management	Virtual team management enables effective working support for people on half-work/half-leisure senior working holidays.
Gamification	Without the legacy ethic of salaried work, gamification provides a day-to-day targets and rewards for the heterogeneous work environment.
Crowd-sourcing	Work items are split into small pieces so that a massive number of small tasks are completed to achieve goals.
Intern education	Youth require team management. With a cross-generational work environment, they can learn how to collaborate with a diversity of people, gaining increased management experience and skills.

Possible social measures that support senior nomadic work are depicted in Table 4.

Table 4. Possible social measures that support nomadic senior work life

Measure	Description
Senior hostels	Turning old residence condos backed by the municipality into short-stay work-life support homes.
Hot spring redevelopment	Turn legacy hot spring resorts into work-life balance stay zones.
Internship programs	Provide summer/winter internship entrepreneurships to the younger generation in the region as short-term virtual working programs
Transition programs	Have enterprise program for age 60-65 to support nomadic working holidays for pre-pension training.
Open innovation infrastructure	Create an open innovation infrastructure to discuss marketing ideas for regional products.
Gamification SaaS	Deploy virtual team gamification as type of Software as a service (SaaS).
Remote medicare	Have time-limited economic zones to provide remote medicare for seniors.

5 Discussion

5.1 The Advantages of the Proposed Approach

The proposed framework provides a basis for addressing not only the social problems emerging from demographic changes, but also provides a new direction for a new type of innovation.

Table 5. Comparison with existing innovation

Existing innovation	Difference
One-directional innovation	A combination that includes multiple value systems, such as age-based and pension-based work lives must be coordinated with the challenges of work and social norms.
Unification innovation	Past innovation focused on simplicity and unification, for the sake of efficiency. The proposed framework involves technology innovation for integrating heterogeneity, diversity and multiplicity of social segments in reality.
Technology innovation	The proposed framework focuses on social multi-generational collaboration and technology innovation just as just a part of it.
Single-generation innovation	The proposed innovation framework deals with multi-generational innovation, which involves non-capitalistic emotion-and-meaning-based collaboration.
Symmetric innovation	The proposed framework serves for heterogeneous social contexts with fundamentally diverse work norms and social norms. They are departures from symmetry.
Quantitative innovation	The proposed framework deals with a heterogeneous and diverse reality in a conflicting multi-generational society without any unified quantitative measures.

5.2 Limitations

This research is descriptive and qualitative. The detailed implementation, deployment, operation, and evaluation of feedback remain for future studies.

6 Conclusion

The existing view of aging is one of decline, decay, lost freedom, life-care and medicare: mostly negative backward-motion. However, the ongoing demographic changes have brought about a need for long-term innovation involving role changes of multiple generations on a massive scale. This innovation requires both bottom-up building blocks and top-down conceptual frameworks.

The author identifies senior human resources as perhaps the last opportunity for innovation in massive-scale value creation. Cross-boundary multi-generational collaboration is a scheme for such an innovation in the 21st century. The framework provided in this paper is a stepping stone for such an opportunity in innovation.

The author proposes a triangular model based on senior working holidays for creating a shift toward a nomadic work-life balance for senior citizens. It provides clues for integrating quality of life for senior citizens, regional industrial development, and on-the-job virtual team management training for youth. The proposed framework provides multi-generational, trans-regional coordination for work-life balance.

References

1. Chesbrough, H.W.: The era of open innovation. In: Mayle, D. (ed.) Managing Innovation and Change, 3rd edn. Sage (2006)
2. Gratton, L.: The shift: The future of work is already here, 1st edn. HarperCollins, UK (2011)
3. Yamakami, T.: Cross-boundary software engineering: Implications of engineering paradigm shift. In: ICIS 2010, pp. 536–540. IEEE Computer Society (June 2010)
4. Yamakami, T.: Virtual-world alchemy: Implications for service engineering from mobile social games in Japan. In: ICACT 2012, pp. 1227–1231. IEEE (February 2012)

A Design Requirement View Model: Lessons from the Evolution of WebKit APIs

Toshihiko Yamakami

ACCESS, Software Solution,
1-10-2 Nakase, Mihama-ku, Chiba-shi, Japan 261-0023
http://www.access-company.com/

Abstract. The design of current OSS requires multi-faceted consideration in order to cope with the requirements of modularity and extensibility. The author discusses the application programming interfaces (APIs) of WebKit and highlights the design requirements of reusable OSS components. From lessons learned, the author proposes a three-dimensional model of the grand design of APIs for reusable OSS components.

1 Introduction

The decrease of information and communication costs has enabled cross-boundary software development such as open source software (OSS). The increasing number of OSS projects has created a new challenge to the design of OSS. OSS has a wide range of design requirements, therefore, it is crucial to coin a sound grand design framework that enables reusable OSS components in the current technology landscape.

WebKit is a web rendering engine toolset. Apple, Nokia, and Google use it for their smartphones, therefore, WebKit is a de facto standard rendering engine for the smartphones. In 2013, conflicts between Apple and Google resulted in the forking of Blink, another OSS project.

The author discusses the design requirements of the WebKit API, the WebKit2 API, and the Chromium API from the viewpoint of abstraction. Then, the author proposes a three-dimensional model for the grand design of an API set of reusable OSS components.

2 Background

2.1 Purpose of Research

The aim of this research is to identify a design methodology for reusable OSS components.

2.2 Related Work

OSS has expanded its coverage in the last two decades as the Internet has penetrated into the software industry. The penetration and variety of OSS has brought about the

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_35, © Springer-Verlag Berlin Heidelberg 2014

need for design disciplines in OSS projects. However, the community computing and voluntary-based computing culture has turned the interest away from any solid design methodologies. Software design is ad hoc and developer-dependent in many cases.

There are three types of existing literature: a) OSS design patterns, b) collaboration in OSS design, and c) Specific design.

With respect to OSS design patterns, Hammouda et al. discussed impacts of the licensing issues on OSS design patterns [4]. Aversano et al. discussed the changes of design patterns in three OSS projects [1]. Stol et al. presented a method for identifying design patterns in OSS projects [7].

With respect to collaboration in OSS design, Zilouchian et al. discussed the consensus building evident in OSS communities [8]. Barcellini et al. discussed cross-participation in design-use mediation [2].

For WebKit-specific design issues, WebKit is the most popular web rendering resource [6]. With respect to the emergence of the WebKit2 API, a high-level description of WebKit2 is described in [5]. Grier et al. used WebKit for their OP2 security-enhanced browser implementation [3].

The originality of this paper lies in its examination of a design methodology for reusable OSS components.

3 API Design Requirements for WebKit

3.1 Retrospective View of WebKit API

The WebKit project is a successful OSS project that enables the development of embedded browsers for a wide range of Non-PC platforms including smartphones. In April 2013, Google announced its fork to Blink. The collaboration between Apple and Google over browsers had ended. This is a good moment to provide a retrospective view of the WebKit APIs. It should be noted that the OSS project name “WebKit” and the API name “WebKit API” for one of the APIs of the WebKit project are different.

3.2 WebKit API

The WebKit OSS project has an API layer “WebKit API”. The WebKit API is a set of APIs that deal with classes such as WebView. The design requirement for the WebKit API is that it works as a part of a platform/library. The API accepts the development environment. Each developer just passes a URL to the WebView class in order to render that URL. It is integrated and easy for each development environment on every platform, such as iOS. The WebKit API has affinity with platforms and development environments. In other words, it does not run without a target platform and development environment. In the case of iOS, theWebKit API is integrated into the Cocoa development environment. The API accepts a Cocoa object. This is same as other development environments such as Qt and GTK.

The relationship between WebKit API and WebKit2 API is depicted in Fig. 1.

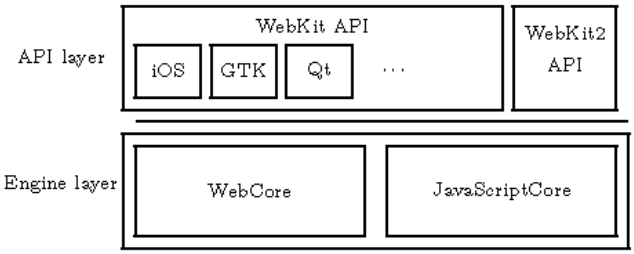


Fig. 1. Relationship between WebKit API and WebKit2 API

3.3 Chromium API

Chromium is an OSS browser project that is supported by Google. Chromium defines the Chromium API which ignoresWebKit API conventions. The Chromium API has an abstraction of platform-dependent functions. ChromiumWebKit API enables multiple-process execution. Therefore, network access through HTTP or local file access is managed by a browser process that has access to the Chromium API for executing WebKit rendering. A browser process manages resource handling on the platform including networking.WebKit is an external process driven by that browser process using inter-process communication.

The Chromium API does not provide any utilities for managing of inter-process communication. Chromium natively deals with the WebCore rendering engine directly without the use of an API. This increases the risk of broken builds when a WebKit project and Chromium project makes asynchronous updates of the source code. Therefore, the Chromium API provides a separation unit that prevents unintentional collisions of source code.

The Chromium API is not intended to be used outside of Chromium at all. Therefore, it is a dedicated API for Chromium.

The inter-process structure of Chromium and WebKit is depicted in Fig. 2.

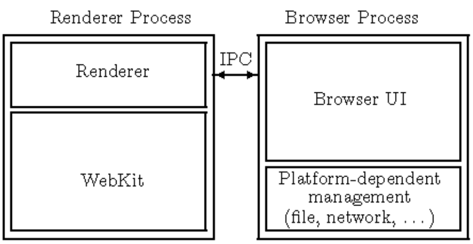


Fig. 2. Inter-process structure of Chromium and WebKit

3.4 WebKit2 API

TheWebKit2 API is another API for theWebKit project. TheWebKit2 API uses the same WebCore and Javascript core as the WebKit API. However, from the viewpoint of an API, there is no compatibility between the WebKit API and the WebKit2 API. It implements an abstract layer that enables a process-separation model in WebKit. The management of inter-process communication is embedded in the WebKit2 API. The WebKit OSS is used in many applications other than a browser. Examples include music players, mail applications and so on. In order to avoid duplicating inter-process communication in each application, the WebKit2 API handles its own management of inter-process communication. This eliminates the reimplementation of the same management functions in each application.

Using the process-separation model, theWebKit2 API minimizes dependency to the platform. A minimum of platform-dependent code is written in platform- dependent code such as Cocoa for iOS, and Qt-C++ for Qt. An example of such platform-dependent code is creating a window. Most APIs are defined in C.

This separation enables the reuse of code. In the WebKit API, people manually copy-and-paste codes for each porting.

The design requirement for the WebKit2 API is the reusability of source code.

The process boundaries of WebKit2 API are depicted in Fig. 3.

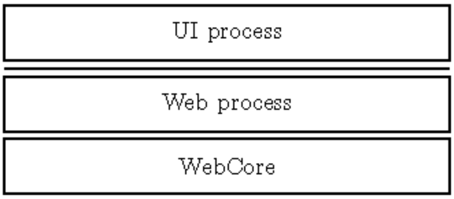


Fig. 3. Process boundaries of WebKit2 API

The design requirements for OSS components are depicted in Table 1.

Table 1. Design requirements

Aspect	Summary
Integration into development environment	Development-friendly APIs makes development easy
Separation of library versioning	Encapsulate from version conflicts with agile development
Process separation	Reuse components as a process-separated entity
Platform-dependency separation	Non-platform-dependent code is clearly separate and maintained
Embedded platform functions	Porting is made easy through the reuse of high-level functions

4 Design Requirements View Model

A view model of component design is depicted in Fig. 4. The three dimensions are programmability, portability, and reusability.

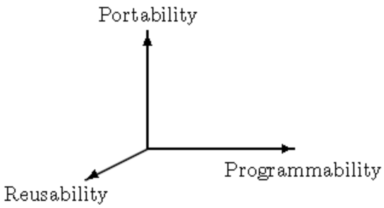


Fig. 4. A view model of component design

The requirements dimension and design decision are depicted in Table 2. Each design decision reflects the tradeoffs between reusable OSS components.

Table 2. Requirements dimension and design decision

Dimension	Design decisions
Programmability	Awareness of control techniques using social competition.
Portability	Porting without modifying for each platform. APIs are designed to maximize the ratio of portable common parts.
Reusability	Multiple applications embedded the component run on multiple platforms. Common management components can be reused across platforms.

Comparisons of every design decision are depicted in Table 3.

The design decisions shift from the top to bottom during the process of spreading among a wide range of diverse platforms.

Table 3. Pros and cons of each dimension

Design decision	Pros	Cons
Tight-coupling of platform	Improves of performance. Easy design decisions with specific requirements	Difficult to port to other platforms. Larger market segmentations.
Development environment friendliness	Easy for programmers. Easy to integrate into a development environment. Easy to integrate platform-specific functions.	Excessive copy-and-paste when porting, leading to poorer quality and increased costs for porting.
Process separation	Strong encapsulation. Strong sandbox. Increased performance with multi-core execution environment. Ease of porting common parts. Strong encapsulation to versioning of co-existing components.	Difficult to manage through direct manipulation (e.g. DOM manipulation in a browser)

5 Discussion

5.1 Advantages of the Proposed Approach

Many OSS projects have its roots in agile, decentralized and community computing, which is a departure from legacy structured software development. The difficulties of formulating a grand design in the initial stages of OSS projects are depicted in Table 4. The proposed view model provides criteria for evaluating the design trade-offs when designing APIs for the multiple design requirements that emerge. The design of APIs has increased in importance with the increased reusability of OSS components.

Table 4. Difficulties of grand design in the initial stages of OSS projects

Aspect	Description
Agile development	Due to the nature of community-development, many programmers prefer coding to design and specifications. This may lead to underestimation of a grand design in OSS projects.
Grass-roots development	Due to the non-paid nature of community-development, many OSS projects rely on developer-based design without having an initial grand design for long-term roadmaps.
Unexpected success	Unexpected success may bring unexpected application domains and unexpected new design requirements.
Time-dimensional factor	Long-lived projects have increased possibility of encountering unexpected technological landscape changes.

The importance of API design for OSS components has recently emerged with a requirement for reusable and stable software components. The designing of APIs is in many cases best-fitted in the platform- dependent cases because they are the most visible and tunable cases. The lessons learned from developing WebKit APIs have shown the importance of having a grand design for the widely-accepted reusable components.

The model facilitates helping both in the initial consideration in the initial stage and in the reconsideration of underlying design assumptions in the adoption stage.

5.2 Limitations

This research is descriptive. Detailed design requirements analysis of multiple OSS projects are beyond the scope of this paper. Political and strategic aspect of OSS design are not covered in this paper.

6 Conclusion

The grass-roots characteristics of OSS have caused the importance of having a grand design in the initial stage of OSS to be underestimated. The more successful an OSS project is, the more conflicting design requirements that emerges. Reusable software components provide APIs that facilitate programming extensibility. The author analyzes the different design requirements of Web components in the case of WebKit.

From a retrospective viewpoint, the author proposes a three-dimensional view model of reusable OSS components. This model facilitates the establishing of a grand design for reusable OSS components in the early stages of their design. It is also usable in the design reconsideration phase after multiple design requirements have emerged.

References

1. Aversano, L., Canfora, G., Cerulo, L., Del Grosso, C., Di Penta, M.: An empirical study on the evolution of design patterns. In: ESEC-FSE 2007, pp. 385–394. ACM, New York (2007)
2. Barcellini, F., Detienne, F., Burkhardt, J.M.: Cross-participants: fostering design-use mediation in an open source software community. In: ECCE 2007, pp. 57–64. ACM, New York (2007)
3. Grier, C., Tang, S., King, S.T.: Designing and implementing the op and op2 web browsers. ACM Trans. Web 5(2), 11:1–11:35 (2011)
4. Hammouda, I., Mikkonen, T., Oksanen, V., Jaaksi, A.: Open source legality patterns: architectural design decisions motivated by legal concerns. In: MindTrek 2010, pp. 207–214. ACM, New York (2010)
5. (2011), The WebKit Open Source Project: WebKit2 at `http://trac.webkit.org/wiki/WebKit2`
6. Raasch, J.: Smashing WebKit, 1st edn. Wiley Publishing (2011)
7. Stol, K.-J., Avgeriou, P., Ali Babar, M.: Design and evaluation of a process for identifying architecture patterns in open source software. In: Crnkovic, I., Gruhn, V., Book, M. (eds.) ECSA 2011. LNCS, vol. 6903, pp. 147–163. Springer, Heidelberg (2011)
8. Zilouchian Moghaddam, R., Bailey, B., Fu, W.T.: Consensus building in open source user interface

Optimizing the Operation Layer Algorithm of NETCONF Protocol

YangMin Lee* and JaeKee Lee

Dept. of Computer Engineering, Dong-A University, Pusan, Korea
manson23@nate.com,
jklee@dau.ac.kr

Abstract. Modern network has a complex structure that includes diverse and heterogeneous equipment to cope with multi-purpose functions. For proper administration of networks with complex structures, NETCONF protocol has been proposed. It still, however, has a few unresolved issues with each layer of the protocol. Tackling one of the less attended issues of NETCONF, this paper suggests a technique which can efficiently update the configuration data of network equipment when multiple administrators are present. Specifically, for efficient updates of the network equipment configuration, we propose to add a pre-examination of the data dependent to <lock> algorithm in operation layer and modify the data structure for managing XML documents for changing the configurations of network equipment. We have implemented and experimented with our method and confirmed that our approach brought a significant overall performance improvement to NETCONF.

Keywords: NETCONF, Operation, heterogeneous, lock, Configuration.

1 Introduction

Modern network, designed to perform various and complex functions, includes multiple heterogeneous equipment and surely has a complex structure. It is essential for this kind of network to have real-time monitoring and rapid response facilities to manage and maintain the performance and functions. To manage a complex network, a large number of network managers and multiple monitoring devices required together with a appropriate protocol. Regarding the network management protocol, existing SNMP showed its limits and NETCONF has been emerged as an alternatives[1]. Currently most network equipment manufacturers are loading NETCONF on their products.

NETCONF is a protocol with 4 modularized layers that have unique functions. It is designed to be suitable for managing complex networks composed of heterogeneous equipment. Being a replacement for SNMP, NETCONF has various functions since it is capable of handling modern network structure. However, it still has some issues concerning the 4 layers since the establishment of the standard[1]. However, NETCONF still has a room for further improvement.

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_36, © Springer-Verlag Berlin Heidelberg 2014

This paper concentrates on improving the algorithmic performance of the Operation layer among the 4 layers. We specifically optimized the operation by adding a data structure to the network equipment configuration information update method [14][15] previous studied and performed a study efficiently improving the <lock> operation [16] by optimizing the data dependency preprocessing technique presupposing a condition in which multiple network administrators are present. Since the performance of NETCONF is increased by improving the algorithm of the Operation layer, we can achieve a performance boost of NETCONF by optimizing the algorithm and network equipment configuration update method. We proved the excellency of our approach by comparing the performances of our improved NETCONF and existing one.

This paper is organized as follows: Section 1 introduces the research goal and backgrounds. In Section 2, we briefly summarized the related work. Section 3 describes our approach to improve and optimize regarding Operation layer. The implementation and comparison of our method and existing method are presented in Section 4. Finally, Section 5 concludes this paper.

2 Related Work

2.1 Studies on NETCONF

There have been many studies on NETCONF, and the results of various attempts to fix the defects of the protocol have been published. For network administration work using NETCONF, YencaP, a NETCONF protocol management system, was created and used in many NETCONF studies [2]. Both in Korea and other countries, a system called XCMS (XML-based Configuration Management System) was also developed and used for analyzing NETCONF protocol and IP sharing equipment compatible with NETCONF protocol [3].

There is a study which developed a network management system consisting of 3 parts including a manager, an agent, and a module group. This system is called NETCONF protocol based BUPT-NET [4].

In a study, a NETCONF protocol based network management system was developed. This study describes the structure of the NETCONF protocol based network management system and operational elements of the system [5]. The manager of NETCONF protocol manages network equipment by transferring monitoring or configuration request messages through Internet. The agent of NETCONF protocol responds to a management request. If a fault occurs, a response regarding the faulty situation is directly transferred to the manager of NETCONF protocol. Objects, in which agents of NETCONF protocol are installed, are managed by NETCONF protocol manager [5]. This system is briefly depicted in Fig. 1.

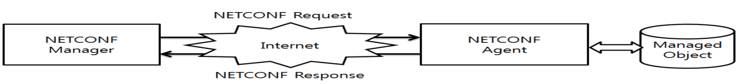


Fig. 1. The structure of NETCONF protocol based network management system

2.2 Studies on Improving the Performance of Operation Layer

Up until now most studies have been focused on techniques for improving performance of the each layer of NETCONF. Among the layers, RPC and Operation layers has been intensively studied for performance enhancement [6][7]. Some studies attempted to reduce the overhead by cutting down the control messages generated in network circuit using NETCONF [8][9][10].

The most popular trend is increasing the efficiency of the algorithm included in Operation layer. A typical approach is enabling the partial lock of a specific memory by modifying the <lock> algorithm [11][12]13][14]. What most existing studies focusing on is solving the problems of standard NETCONF and enhancing the performance. In general, NETCONF with improved algorithm in RPC layer or Operation layer can perform better in processing speed and other features than standard NETCONF [15].

3 Proposed Protocol

We suggest two methods for solving problems found in considering previous work. First, we introduce a data dependency pre-examination technique to add a function that can partially lock the data store often causes a trouble when multiple administrators access single network equipment. Second, we suggest an improved XML document update method for efficiently updating network equipment configuration information when multiple administrators access single equipment.

3.1 Data Dependency Pre-examination Method Using XPath

We basically reused the method suggested by previous work [16] to improve the <lock> algorithm of NETCONF. This is a method of grouping the data with dependencies before updating network equipment configuration information. With this method, we can prevent <lock> locks the whole data store in later steps. When using NETCONF, network equipment configuration information is constructed based on XML document and it is possible to group the data upon the dependencies using XPath in advance.

With XPath, we search for other nodes from current node. We can grasp the relationship between current node and other data nodes using the relation and position represented in XPath. This paper uses all available relation information and attributes values for data grouping. An example is given in Fig.2. This example shows the case of searching for nodes where the port attribute is set 1234. The target node can be found regardless of the position of the port.

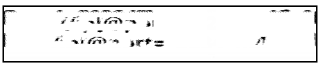


Fig. 2. Example of XPath

By using this kind of operation and applying location steps including child and wine and predicates that can be specified in only selected nodes, we can separate and group the nodes. For instance, we can separate and group the related data in XML document in which node data is serialized as shown in Fig. 3.

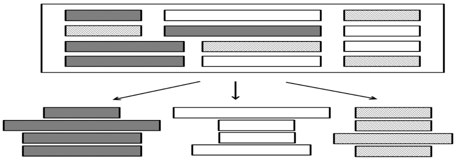


Fig. 3. Example of data dependency examination and grouping

When the examination of data dependency is completed, it is possible to structure the management data belong to the network administrator currently working on into grouped data set suitable for each administrator. From the data being accessed by an administrator other items linked vertically are grouped as a single data set and items linked horizontally in sibling relation can be separated as an independent data set. The separated data set is used in <lock> algorithm in Operation layer. Consequently it is possible to implement an improved partial lock algorithm like <lock-p> [15][16].

3.2 Improving Equipment Configuration Information Update Method Using Queue and List

If administrators make sequential access when updating network equipment configuration, equipment configuration information update method is not required. However, when multiple administrators access a single network equipment in a network consisting of heterogeneous equipment, strategic information update method is required. Generally used method is to maintain multiple candidate documents for running-configuration modification. In this method, running-configuration update data needs to be maintained and managed as multiple XML documents in accordance with the number of administrators [13]. Maintaining a XML document for each of the multiple administrators and processing them sequentially are somewhat inefficient.

We created a rule set for updating XML candidate documents for administration. By doing so, we give an order of priority to administrators according to their levels and privileges and grant higher priority to the instructions that have wider range of change to the network equipment among the administrates possessing a same level of privileges [16].

As shown in Fig. 4, administrators' XML documents are stored sequentially in a general queue. After going through the candidate checking module, the candidate documents can update the running-configuration of network equipment. XML documents are dispatched into a list in the order of passing through the previously mentioned update rules.

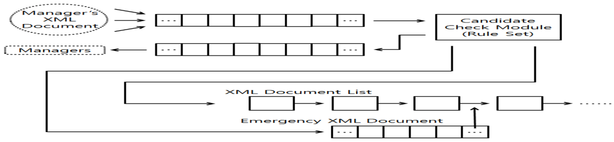


Fig. 4. System data structure for XML candidate document processing

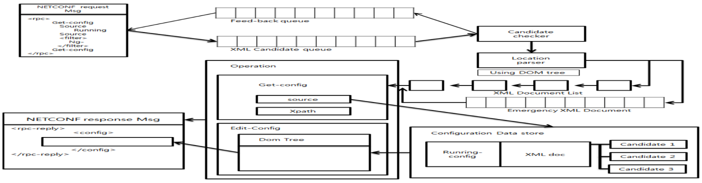


Fig. 5. NETCONF with applications of proposed method

The operation of the method proposed in this article can be briefly illustrated as in Fig. 5 focusing on NETCONF agent. For an immediate feed-back to the manager, we use a pair of queues. Candidate documents are stored in the XML document list after suggested update rules are applied and checked passing through the location parser to identify the network equipment configuration to be updated. XML documents carrying urgency flag are stored in the urgent XML document queue and moved in front of the XML document list.

4 Experiment and Analysis

4.1 Experiment Environment

We used C language for implementation and performed experiments in a local environment consisting of a single computer. The computer used in our experiment is equipped with an Intel i7 processor and 4GB memory. We used Windows operating system. The numbers of administrators were limited to 10. One of them was a top-level administrator and they were given a same level of privileges. The processing time of using XPath was measured utilizing a Windows system function. We quantified the result of comparing the NETCONF method and our method using an algorithm. Comparisons factors are processing time, network function maintain probability, instruction response speed, number of generated control packets.

4.2 Comparison of XPath Processing Time and Network Function Maintain Probability

XPath processing time is time taken for date dependency pre-examination for creating <lock-p> algorithm in Operation layer. Since existing NETCONF cannot use partial <lock>, we need to measure the time cost and CPU consumption even though pre-examination of data dependency using XPath is required. In our experiment, we measured the time taken from the point of XML document input to the point of applying the result in existing NETCONF. This is equivalent to measuring the time taken by grouping data when XPath is executed. Fig. 6 outlines the comparison factors used in our experiment.

According to Fig. 7, the average XML document processing time of existing NETCONF is 0.21 second faster than that of the method applying XPath. The results were obtained when we limit the size of the documents as 1 Kbyte, and we can consider this time as the time needed for XPath operations. However, the CPU consumptions for two methods are less than 2%. Although using XPath consumed additional time, the network function maintains probability has been increased by using <lock-p> (Fig. 8).

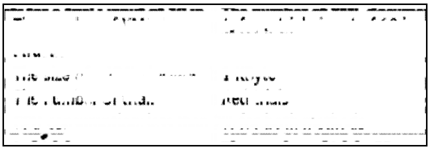


Fig. 6. Outlines the comparison factors (left)

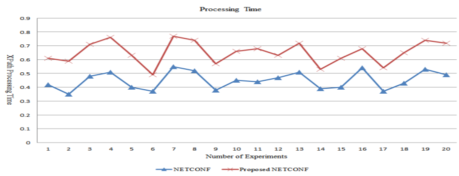


Fig. 7. XPath processing time and comparison of existing NETCONF (right)

In existing NETCONF, a large amount of data is transferred when executing <get-config> and <get> algorithms retrieving equipment configuration data after data dependency examination using XPath. Occasionally, network functions are partially paralyzed because <get-config> algorithm ignores the status data of network equipment in certain cases, e.g. when the port configuration for a router is wrong. Similar incident can happen when the whole memory of network equipment is updated after the <lock> algorithm locks the whole memory and the configuration data is updated. The network function maintains probability is shown in Fig 8. In proposed NETCONF, <lock-p> algorithm has little effect on the configuration data of an equipment as it is able to perform a partial <lock> algorithm using XPath for dependency examination. This can be quantified as follows Fig. 9:

The dependency pre-examination suggested in this paper is clearly slow because this method uses XPath. However, when we assume the maximum number of equipment as 50 and the number of administrators as 10, the router halt probability is less than 3% as <lock-p> algorithm can be used which partially locks data store through dependency pre-examination. In existing NETCONF, router halt probability is more than 9% since the whole data store is locked when network administrators give duplicate instructions or access identical equipment. The <get-config-m> does not actually ignore equipment status data pre-executing the comparison of the status data and configuration data[15]. Thus, proposed NETCONF has approximately 8.8% higher network maintain probability than existing NETCONF.

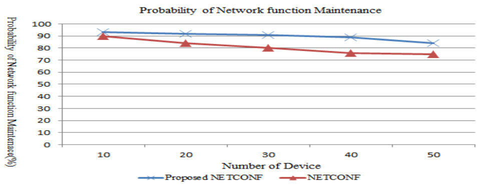


Fig. 8. Network function maintain probability

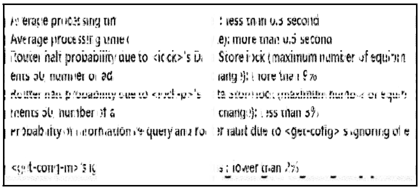


Fig. 9. Quantified factor for XPath processing time

4.3 Comparison of Network Response Speed and Overhead via Improving Equipment Configuration Data Update Method

This experiment is concerned about the method for efficiently updating equipment configuration data when we want to update configuration data of a single machine assuming multiple administrators' presence. When multiple administers simultaneously access the memory which stores configuration data and give certain instructions, the operation result should be applied as soon as possible and the administrators can check out the results.

Fig. 10 shows the result of comparing the instruction response times for the increase of the number of administrators. The result reveals that our NETCONF's average instruction response speed is 0.8 second faster than existing NETCONF regardless of the increase of the number of administrators. The evidential factors for the result are shown Fig. 11.

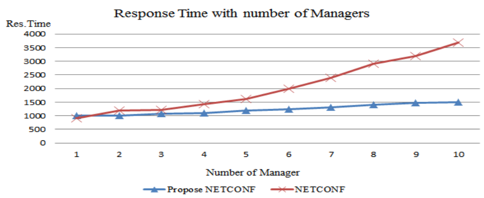


Fig. 10. The Instruction Response Time for the Increasing number of Administrators (left)

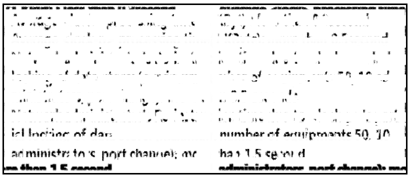


Fig. 11. Evidential factors for the Response Time (right)

The analytical basis of this experiment is that proposed method needs internal processing time for examination using <lock-p> algorithm for data dependency pre-examination. In addition, there is a burden of using 2 queues and 1 list for network administrator's efficient management of the equipment configuration update information. Generally, however, data transfer time through network consumes far more time than internal processing. As shown in the presented values, the difference between operation processing times of <lock> and <lock-p> can be safely ignorable when comparing the time taken for reissuing instructions due to the whole locking of the data store and the response times. Consequently, the response speed of our NETCONF is faster than existing NETCONF since multiple candidates for XML documents can exist and network administrators can update equipment configuration data in parallel with our NETCONF while administrators are required to wait for other administrator's job is done in existing NETCONF.

Finally, we performed an experiment to check out the excessive overhead is created by the control packets generated to obtain faster response speed. If excessive overhead is created following the increase of the number of administrators, there is not much meaning for small speed improvement of the response time.

Fig. 12 shows the tendency of the increase of the number of control packets for increasing number of administrators. When the experiment is performed under an assumption that each administrator regularly generates a packet containing an instruction, the number of generating control packets in network increases in existing NETCONF to get confirmation and response since some administrator's instructions are not executed. This tendency is getting stronger as the number of administrators increases.

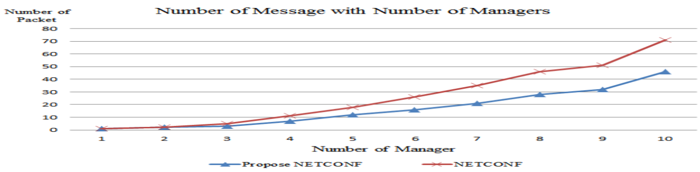


Fig. 12. The number of control packets for increasing number of administrators

When the network equipment configuration update method proposed in this article is applied, administrators can check the reasons for unexecuted instructions by getting rapid feed-backs. The number of control packets does not greatly increase as network equipment configuration updates can be processed in parallel although the number of administrators increases.

Proposed NETCONF shows an excellent performance regarding the number of control packets transferred in the network in Fig. 12. Significant performance boost is observed only when the number of administrators is more than 7.

5 Conclusion and Future Work

Modern network is designed for performing multi-purpose functions and consist of heterogeneous equipment with a complex structure. Consequently, monitoring and malfunction management became very important to maintain the performance of a network requiring huge time and human resources including multiple administrators.

NETCONF was proposed for efficient management of complex networks. However it still has unresolved issues in each layer of the protocol.

This article suggested a network equipment configuration data update method when multiple administrators are present among the issues. This topic did not get much attention in other studies. We improved the efficiency of <lock> algorithm in Operation layer using data dependency pre-examination. We also achieved a better performance regarding efficient equipment configuration update by modifying the data structure for managing XML documents for network equipment configuration. We confirmed that our improved NETCONF has 8.8% of higher network function maintain probability than existing NETCONF by implementing the algorithm based on queue theory and testing it. The instruction response speed is 0.8 second faster than existing NETCONF. By contrast, the numbers of generated control packets are less than existing NETCONF.

The future work is confirming the practicality and efficiency of our algorithm by applying it to NMS for real network management.

Acknowledgment. This paper perform that funded by the government in 2014 (Ministry of Education, Science and Technology) and supported by the National Research Foundation of Korea basic research project. (2013-054590)

References

1. Enns, R. (ed.): NETCONF Configuration Protocol. RFC 4741 (2006)
2. Yang, `http://www.netconfcentral.org/static/slides/yangtutorial/yang_getting_s24tartd2.html`
3. Choi, M.-J., et al.: XML-based Configuration Management for IP Network Devices. IEEE Communications Magazine 41(7), 84–91 (2004)
4. Huang, J., Zhang, B., Li, G., Gao, X., Li, Y.: Challenges to the New Network Management Protocol-NETCONF. In: First International Workshop on Education Technology and Computer Science, ETCS 2009, pp. 832–836 (2009)
5. Chang, Y., Xiao, D.: Design and Implementation of NETCONF-Based Network Management System, pp. 256–259. IEEE Computer Society (2008), doi:10.1109/FGCN
6. Yoo, S.-M.: Performance Improvement Methods for NETCONF-Based Configuration Management. Computer Science, 242–252 (2006)
7. Ji, H.: Challenges to the New Network Management Protocol-NETCONF. In: First International Workshop on Education Technology and Computer Science, ETCS 2009, pp. 832–836 (2009)
8. Chang, Y., Xiao, D.: Design and Implementation of NETCONF-Based Network Management System. In: 2008 Second International Conference on Future Generation Communication and Networking, IEEE Computer Society Magazine, pp. 256–259 (2008)
9. Schonwalder, J., Bjorklund, M.: Network configuration management using NETCONF and YANG. IEEE Communications Magazine 48(9), 166–173 (2010)
10. Yu, J., Al, I.: An Empirical Study of the NETCONF Protocol. IEEE Computer Society Magazine, 253–258 (2010)
11. Lee, M.: Improved Performance of network configuration management and System base on NETCONF. Korea Communication Science Society Journal 33(9), 787–790 (2008)
12. Apostolos, E.: Management Traffic in Emerging Remote Configuration Mechanisms for Residential Gateways and Home Devices. IEEE Communications Magazine 43(5), 154–162 (2005)
13. Mi-Yang, C., Lee, Y.-M., Lee, J.-K.: NETCONF Protocol with Improvements Renewal Method for Device Configuration Data. 2013 Korean Society for Internet Information Spring Conference Magazine 14(1) (2013)
14. Lee, Y.-M., Lee, J.-K.: Development Renewal Method for Device Configuration Data of NETCONF Protocol considered Multiple Network Managers. In: IEEK Summer Conference 2013. Paper CD (2013)

15. Cha, M.-Y., Won, J.-S., Lee, J.-K.: Network Circuit Auto Recovery using an Improved NETCONF Protocol. Journal of the Korean Institute of Information Scientists and Engineers 18(9), 628–638 (2012)
16. Lee, Y.-M., Cha, M.-Y., Lee, J.-K.: Development of Update Methods for Configuration Data of NETCONF Protocol considering Multiple Network Administrators. Journal of Korean Society for Internet Information 14(5) (2013)

Research about Virtualization of ARM-Based Mobile Smart Devices*

Lei Xu, Wenzhi Chen, and Zonghui Wang

College of Computer Science and Technology
Zhejiang University, Hangzhou, P.R. China
{leixu,chenwz,zjuzhwang}@zju.edu.cn

Abstract. In common sense, virtualization technology is adopted to offer several isolated execution environments and makes better use of computational resources which has been an important enabler for cloud computing. However, in embedded systems, the significance of virtualization does not come into the picture. The extensive utilization of mobile smart devices has led to a series of issues such as security, power consumption and performance limitation. Mobile virtualization can offer an effective approach in addressing these challenges. In this paper, we discuss how mobile virtualization addresses these challenges and then present a detail analysis of mainstream mobile virtualization solutions: ***Para-virtualization***, ***Hardware-Assisted Full virtualization*** and ***Microkernel Hypervisor***. At last, we carry out a series of performance comparison between these solutions and make some suggestions for further research.

Keywords: Smart Devices, Mobile Virtualization, ARM, Android.

1 Introduction

The ability of traditional virtualization brings immense benefits in terms of reliability, efficiency and scalability. It enables the datacenters to flexibly provision resource which makes the *computing-as-a-service* vision of cloud computing possible [1]. A substantial amount of works have been carried out on traditional virtualization most of whose architecture is X86.

Nevertheless, ARM-Based mobile smart devices are becoming more and more ubiquitous and the preferred platform for users' daily computing needs are shifting from traditional desktop to mobile smart devices [2]. Undoubtedly, as mobile computing advances, it brings several tough challenges, as described follows:

- **Security Threats.** Mobile device, as a kind of intimate personal portable equipment, contains lots of user's sensitive data, such as SMS, contacts and photos. People can't pay much more attention on its security issues, especially in a poor secure condition nowadays.

* This research is funded by National Science and Technology Major Project of the Ministry of Science and Technology of China under grant NO. 2013ZX03003010-002.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_37, © Springer-Verlag Berlin Heidelberg 2014

- **Performance Excess.** Multi-Core SOC is increasingly adopted by hardware vendor along with 2G RAM or more. It seems that these vendors are participated in a hardware competition which led to a serious performance wasting. How to make better use of these multi-core hardware resources is a new challenge.
- **Power Consumption.** Power is always the bottleneck of mobile devices. Especially, modern device architecture is becoming more and more complicated to support various modem protocol stacks (GSM, WCDMA, and LTE) simultaneously and many complex applications. People want to find a way to simplify the hardware architecture.
- **Shorter time-to-market.** For devices manufacturer, they wish to quick release their newest products to meet dynamic market requirements. They want to find a way to reduce dependencies among hardware and software components so as to reuse legacy software or legacy operating system on a new design chip/board and reduce development and integration time and effort.

To address these challenges, the role of virtualization within the mobile device is being discussed among academia and industry [3]. Actually, mobile virtualization can deal well with these challenges. But this technology seems has not yet aroused people's enough attention until now. So our team carried out this research.

In this paper, we comprehensively analyze mobile virtualization technology: Section 2 describes the definition and the benefits of mobile virtualization. Section 3 discusses the mainstream solutions in detail. Section 4 carries out performance comparison among those solutions. A summary is described in Section 5.

2 Mobile Virtualization Overview

Mobile device, a modern embedded system, is increasingly taking on characteristics of general-purpose systems. Their functionality is growing, and so is the complexity of their software [4]. This creates a demand that run more and more high-level applications originally developed for the PC world, such as virtual machines. Mobile virtualization is a variant of system virtualization that enables multiple isolated OSs run simultaneously on a single mobile device.

However, mobile device is a personal communication device rather than a totally computing devices. This means that we can't deal with it like PC which is obviously not adequate to mobile virtualization [5]. The requirements of mobile virtualization includes: (1) A small code size and lightweight hypervisor; (2) A strict system-wide security policy; (3) Strong interaction to enhance user experience; (4) Minimal impact on system resources and real-time performance, and so on.

2.1 Mobile Virtualization Benefits

2.1.1 Enhanced Security

The security issues of mobile smart devices are heavily exposed [6]. Viruses, Trojan horses and malwares from all kinds of external attackers have caused people's attention. However, deploying a security environment (such as encryption, digital signature, safety audit, access control, digital certification, *etc.*) on mobile device is very hard for common users. So people need an innovative solution which can offer a secure and credible execution environment when use some critical applications (mobile banking), or access to sensitive data (SMS, contacts).

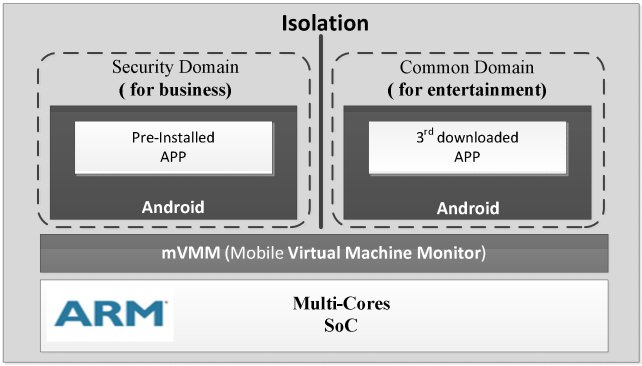


Fig. 1. Multi-OS isolation for enhanced security

Mobile virtualization is such a kind of solution! As shown in Fig.1, a security domain contains pre-installed application for the basic functions of a mobile, such as SMS, dialer and mailer. The 3rd-party downloaded applications can never affect the base domain, which are only allowed to execute in a common domain. The isolation offered by the mobile virtualization makes this possible. We can use security domain for private telephony, business office, mobile banking and so on. Also, we can create several common domains for daily use (browsing, gaming, movie, music, *etc.*).

2.1.2 Simplify Hardware Structure

Current device architecture is very complicated and inevitably brings power consumption problem, as shown in Fig.2. Each core has the different purpose: ARM-A runs general-purpose operating system, which is up to interact with users; ARM-C runs a real time operating system, mainly to complete high-level protocol stack processing of different communication formats; DSP-C has strict real-time requirements, mainly to process underlying protocol stack by interrupt trigger; while DSP-M is always used to decode audio and video.

Mobile virtualization breaks the tightly one-to-one relationship between operating systems and processors. How to enable devices support multiple new protocol stacks (GSM/WCDMA/HSPA/LTE) and avoid compatibility problems between different protocol stacks is a challenge. Instead of using multiple dedicated real-time processors, mobile virtualization offers a new architecture as shown in Fig.3. In this architecture, VMM supports multiple general purpose operating systems (GPOS) and real time operating systems (RTOS) to run concurrently on one ARM processor. And some underlying protocol processes can be scheduled to the unique DSP by the VMM. This architecture can authentically simplify hardware structure.

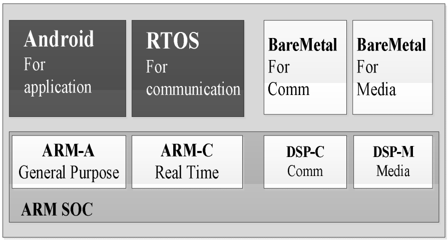


Fig. 2. Typical mobile device architecture

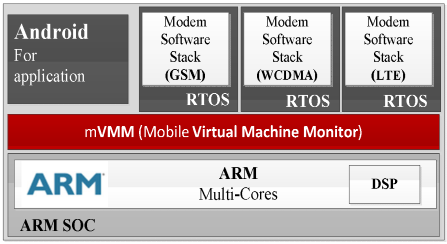


Fig. 3. Simplified hardware architecture

2.1.3 Reuse Legacy Software

Mobile virtualization empowers mobile device manufacturers and semi-conductor vendors to speed time to market and reduce costs by reusing legacy software assets while taking advantage of new designed chip. Maintaining a competitive edge is vital for mobile device manufacturers, who must integrate huge amounts of complex software on multiple chipsets and hardware platforms. Native or proprietary device drivers, protocol stacks and system modules can be integrated with ease, and legacy applications can run unmodified in the new environment because of mobile virtualization. This ensures minimum development cost and faster time to market for new products.

3 Mainstream Mobile Virtualization Solutions

3.1 Solution 1: Para-virtualization of ARM

Para-virtualization is a very mature technology used by Xen, a famous hypervisor. It refers to a technique where the guest operating system is modified and privileged instructions are replaced with calls to the hypervisor named *hypercalls*. The hypervisor layer provides a *hypercall* interface with services such as memory management, device usage and interrupts management to the guest. This ensures that all privileged mode activities are moved from the guest operating system to the hypervisor. Since para-virtualization requires changes to the guest operating system code to avoid calls to privileged instructions, it obviates the need for trap & emulate and binary translation. Of course, this benefit comes with the additional cost of maintaining a modified guest operating system.

This solution offers several advantages:

- **Relatively High Performance:** Para-virtualization provides specially defined hooks to allow the guest(s) and host to request and acknowledge tasks, which would otherwise be executed in the virtual domain, so it can reduce the portion of the guest's execution time spent performing operations.

There are some drawbacks:

- **Poor User Experience:** Solutions based on para-virtualization is not fit for mobile device. It has a complex configuration which is not easy for common user and it need to modify the guest OS code which means it can't support the latest OS and commercial closed-source operating systems.

3.2 Solution 2: Hardware-Assisted Full Virtualization of ARM

ARM announced their Cortex-A15 processor with architectural virtualization support in 2010. This virtualization extension provides a new processor mode (HYP mode) and a number of features to improve performance [7]. HYP mode is entered from other modes via a new instruction (hvc), and optionally on a configurable set of exceptions from user or kernel mode. It has banked registers, as well as additional hyp-only registers for system configuration and information on the event which caused entry of hyp mode. There is a hyp-only virtual machine identifier (VMID) register. TLB entries are tagged with the VMID, which supports coexistence of mappings from multiple guests and thus eliminates the need to flush the TLB on a world switch.

This solution offers several advantages:

- **Hardware Support:** This solution is the exclusive way that makes use of ARM hardware feature. This means it can reduce code size and increase reliability. Predictably, this solution will be the major way used in ARM-based machines, even will be applied on the ARM-based server.

There are some drawbacks:

- **Not Mature Enough:** As mentioned above, ARM virtualization extension is a new technology. Meanwhile, KVM is not originally designed based on ARM architecture. So this solution is now not mature and not stable, there is a long way to modify KVM to be adaptable to ARM hardware extensions.
- **Poor I/O Virtualization Ability:** As we know, KVM leverages QEMU (an open-source hosted hardware emulator) to offer the ability to virtualize diverse I/O devices, which is a so heavyweight software that not suitable for mobile devices. So this solution has the poor ability to virtualize I/O device without the support of QEMU.

3.3 Solution 3: Microkernel Hypervisor

With virtual machine based on microkernel architecture, we can convert hardware resources to various real-time system services, and deliver to client operating systems which run on virtual machine by mode of virtual devices. In this way, it can support real-time and non-real-time applications to run simultaneously, and provide a universal and transparent interactive interface between non-real-time applications and real-time system functions. The microkernel approach leads to a system structure that differs significantly from that of classical operating system.

This solution offers several advantages:

- **Efficient Resource Sharing:** Microvisor provides mechanisms for efficient sharing of resources. Arbitrary memory regions can be shared by setting up mappings between address spaces (providing high-bandwidth communication channels).
- **Flexible Scheduling:** Microvisor allows the guest operating system to select the appropriate global scheduling priority which means it can run at a high priority when executing real-time threads, and a lower priority when executing background tasks.

There are some drawbacks:

- **Device Emulation:** Microvisor has to provide device support and emulation, an onerous requirement for mobile devices which provide increasingly diverse hardware devices. For example, we are not aware of any OKL4 implementations that run Android on any phones other than the dated HTC G1.

4 Performance Comparison

We have carried out a series of experiments to evaluate the performance of these different solutions described in Section 3. To make a comparison, we choose out their mutual features to test. We built three open source platforms representing correspondingly those solutions: ***CodeZero*** (Microvisor), ***KVM-ARM*** (Hardware-Assisted Full virtualization) and ***EmbeddedXEN*** (Para-virtualization). Meanwhile, we use Urbetter S5PV210 development board with Exynos 4412 CPU and 2GB memory as our experiment platform. In addition, our host OS is Archlinux with 3.5.4 kernel and our guest OS is Android 4.2.2. At last, we choose LMbench3 to be our benchmark.

4.1 Evaluation Results

4.1.1 Context Switching

We measure context switching time between guest OS and host OS. A context switch is the switching of the CPU from one process or thread to another. When the VMM receives a hardware interrupt, it generally suspends the progression of the current process and starts servicing the interrupt. This is an important feature for mobile software which means a good user experience. As shown in Table 1, hardware-assisted solution has the fastest switch speed.

Table 1. Context Switching Time

	Full Virtualization	Para Virtualization	Microvisor
Average Time(μs)	18.3	30.1	23.7

4.1.2 Micro Benchmarks

We used the LMbench benchmark suite to evaluate the performance of *fork()+exec()*, *fork()+exit()*, pipe and syscall. As shown in Table 2, the performance of executing a simple syscall is the most severely impacted because its execution path is very simple. The other benchmark programs involve fair amounts of work that is executed in the guest OS, thus the performance degradation is a little severe. Among them, Para-virtualization solution got a relatively good result.

Table 2. Preliminary Performance

	Full Virtualization	Para Virtualization	Microkernel
fork + exit (μs)	4,328.53	4012.38	5,117.75
fork + exec (μs)	6,211.51	5,984.14	7,463.90
pipe (μs)	173.30	201.64	1,190.35
syscall (μs)	17.21	13.74	19.93

4.1.3 Macro Benchmarks

In order to see the virtualization's performance impact on common operations in mobile phones, we compared UI loading time, codec performance and image file saving time in table 3. For UI loading test, we used Qtopia installed at NOR flash memory. We prepare 100 files whose size are distributed from 10KB to 5MB to test image file saving and we measure the time taken to save all those image files from a NFS server to NAND flash disk. For codec tests, WMV stream encoder/decoder is used.

Table 3. UI Performance Evaluation

	Full Virtualization	Para Virtualization	Microkernel
UI loading (s)	12.32	13.45	5,117.75
Image saving (s)	45.17	54.23	7,463.90
Encoding rate (fps)	5.67	4.76	1,190.35
Decoding rate (fps)	20.41	23.14	19.93

4.1.4 Scalability Analysis

To analyze the scalability performance impact of the number of concurrent VMs (n), we tested four cases: n =1, 2, 3 … 20 for iterated 10 times shown in Fig.4. Root filesystems are mounted as read-only, then we run a daemon process simultaneously on all running VMs to calculate CPU utilization. Average throughputs values of domains are aggregate throughputs don't degrade much as n increases.

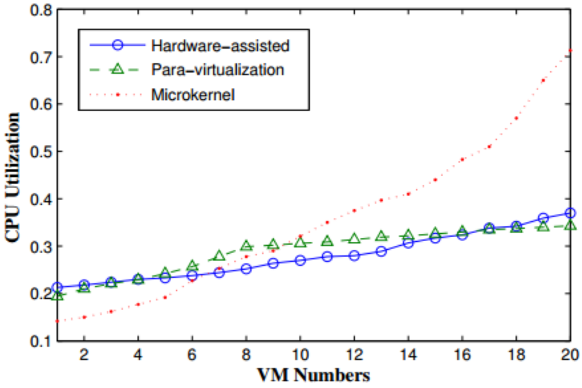


Fig. 4. Scalability performance evaluation

5 Conclusions

Virtualization of mobile devices is becoming a hot research point. Many IT companies such as VMware, OK Labs, Samsung and Red Bend have changed their attentions on this field. In this paper, we described what mobile virtualization is and the benefits it brings. We introduced these solutions in detail and talked about their advantages and limitations. At last, we built an experiment platform and carried out a series of performance evaluation among three open source projects. In the future, we plan to explore several lightweight mobile virtualization solutions, like container technology and multi-execution environments based on detached filesystems.

References

1. Xu, L., Chen, W., Wang, Z., Yang, S.: Smart-DRS: A strategy of dynamic resource scheduling in cloud data center. In: 2012 IEEE International Conference on Cluster Computing Workshops (CLUSTERWORKSHOPS), pp. 120–127. IEEE (2012)
2. Andrus, J., Dall, C., Hof, A.V., Laadan, O., Nieh, J.: Cells: a virtual mobile smartphone architecture. In: Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, pp. 173–187. ACM (2011)
3. Yoo, S., Liu, Y., Hong, C.H., Yoo, C., Zhang, Y.: Mobivmm: a virtual machinemonitor for mobile phones. In: Proceedings of the First Workshop on Virtualizationin Mobile Computing, pp. 1–5. ACM (2008)
4. Heiser, G.: The role of virtualization in embedded systems. In: Proceedings of the1st Workshop on Isolation and Integration in Embedded Systems, pp. 11–16. ACM (2008)
5. Chen, X.: Smartphone virtualization: Status and challenges. In: 2011 International Conference on Electronics, Communications and Control (ICECC), pp. 2834–2839. IEEE (2011)
6. Kizza, J.M.: Mobile systems and their intractable social, ethical and security issues. In: Ethical and Social Issues in the Information Age, pp. 281–297. Springer (2013)
7. Dall, C., Nieh, J.: Kvm for arm. In: Proceedings of the Ottawa Linux Symposium, Ottawa, Canada (2010)

A Hybrid Multi-Criteria Decision Support Model: Combining DANP with MDS

Chen-Shu Wang[1] and Shiang-Lin Lin[2]

[1] Graduate Institute of Information Management,
National Taipei University of Technology, Taipei, Taiwan
wangcs@ntut.edu.tw
[2] Department of Management Information Systems,
National Chengchi University, Taipei, Taiwan
shiang0623@gmail.com

Abstract. Human beings encounter the problem of making decision in their daily life. However, most decision makers, when encounter the decision problems involving multi-criteria or diverse alternatives, they could not make correct decision due to lacking related decision-making information. Therefore, what is most concerned by the decision makers is how to pick out the most optimal decision-making evaluation factors and the best execution alternataive for a multi-criteria decision making (MCDM) problem. This study combined the merits of various decision-making analytic methods, namely, decision making trial and evaluation laboratory (DEMATEL), analytic network process (ANP) and Multidimensional Scaling (MDS), so as to propose a four-stage Hybrid Decision-Making Support Model (HDMSM) to assist the decision makers in making the best decision when they face a decision problem.

Keywords: Multi-Criteria Decision Making, Analytic Network Process, Decision Making Trial and Evaluation Laboratory, Multidimensional Scaling.

1 Introduction

Human beings encounter the problem of making decision all the time in their daily life. Most of these decision problems involve many factors that are to be evaluated by a decision maker, and there are usually many alternatives for one decision problem (Simon, 1977). However, most decision makers, when encounter the decision problems involving multiple evaluation factors or alternatives, often fail to make correct decision due to lacking related information (Hwang & Yoon, 1981). Therefore, what is most concerned by the decision makers is how to pick out the most optimal decision-making evaluation factors and further find out the best execution alternataive for a multi-criteria decision making problem (Yoon & Hwang, 1985).

In the field of decision science, multi-criteria decision making (MCDM) methods are most frequently used to solve the above-mentioned decision problems (Belton, 1990). The objective of MCDM methods is to help the decision makers express their preferences structure from limited number of possible alternatives, and then use

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_38, © Springer-Verlag Berlin Heidelberg 2014

various multi-criteria decision-making analytical methods, such as DEMATEL, AHP and so on, to convert the qualitative value of each evaluation criterion and alternative into quantitative weight (Buede & Maxwell, 1995), and finally, according to the priority ranking, determine the important decision-influencing factors and the ideal execution alternative (Opricovic & Tzeng, 2004). MCDM methods have been rapidly developed in the fields of management and social science in recent years, and have been widely employed in, for example, investment portfolio (Ehrgott, et al., 2004), supplier selection (Shyur & Shih, 2006), and green supply chain (Büyüközkan, 2012).

While there are numerous researches and applications of MCDM methods. However, these methods still have several disadvantages need to be improved. First, each of the MCDM methods has its own theoritical basis as well as its merits and drawbacks. As a result, when different methods are used for the same decision problem, they would usually lead to different results, so that the decision makers are at loose ends (Keeney, 1992). Therefore, it is very important that how to combine a veriety of MCDM methods in order to develop a decision support model for effectively assisting the decision makers in making a correct decision (Yang, et al., 2008). Second, the diversified social environment makes the decision problem much more complicate. In the current environment of decision problem, the evaluation factors frequently have interaction or conflict with each other, and the conventional hierarchy-based MCDM methods just could not accurately help the decision makers evaluate the factors in such decision problems (Saaty & Vincke, 1988).

In view of the above fact, ANP is one of the important MCDM methods that used in the multi-criteria decision making to effectively handle the dependence and feedback among different evaluation factors (Saaty, 2001). Before using ANP to measure the weights of the evaluation factors, a networking among all factors must first be established (Wu, 2008). However, the establishment of the networking is not included in the scope of ANP. Therefore, it need other analytical methods to find out the relation between different factors (Yang & Tzeng, 2011).

ANP can help the decision makers to find out what are the critical evaluation factors in the decision problem; DEMATEL can establish the relation between the evaluation factors; and MDS can be used in multi-alternative decision to analyze the similarity between different alternatives and further assists the decision makers to more accurately find out the most optimal alternative (Huang, 2005). The objective of this study is to combine the merits of the above three decision-making methods, so as to propose a Hybrid Decision-Making Support Model (HDMSM) to assist the decision makers in making the best decision when they face a decision problem.

2 Literature Review

2.1 Decision Making Trial and Evaluation Laboratory

Decision making trial and evaluation laboratory (DEMATEL) was originated from the Geneva of the Battelle Memorial Institute in 1973. It can effectively observe the level of mutual influence among different factors, so as to understand the complicated cause-and-effect relationship in the decision problem Fontela & Gabus (1976). The analytic process are shown as follow.

Define the Relationship among Evaluation Factors. Through literature review or brainstorming, then, listing the factors which affect the problem of decision-making, and interview the experts who are in the related field, in order to determine the relationship between each of two factors.

Establish Direct-Relation Matrix. If the decision problem with n evaluated factors, according to the degree of influence scores which is determined by experts, further, to establish an $n*n$ direct-relation matrix, which represent as Z. Among the matrix, z_{ij} represent the degree of the factor z_i effect factor z_j. The calculation is in equation (1).

$$Z = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \begin{bmatrix} 0 & z_{12} & \dots & z_{1n} \\ z_{21} & 0 & \dots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & 0 \end{bmatrix} \end{array} \quad (1)$$

2.2 Analytic Network Process

ANP is a decision-making analytical method that uses network and nonlinear structure to represent a decision problem, and is developed in response to the fact that many decision problems could not be presented with the structured hierarchy. The main objective of ANP is to correct the traditional AHP, with which the problems of dependence and feedback might occur between the criteria or the layers (Saaty, 1996).

ANP mainly through Supermatrix to show the relationship and strength of graph among factors. The best advantage of using Supermatrix is it can evaluate the external and internal of dimensions dependability efficiently. In addition, we can obtain the weight of each factors through the Limit Supermatrix.

2.3 Multidimensional Scaling

Multidimensional scaling (MDS) is a data reduction method, it uses the distance or similarity between data points to locate the spatial coordinates and the relative positions of several given data in the low-dimensional space (Torgerson, 1952).

M-MDS is mainly through compute the Euclidean distance between each two factors, and show all factors in Perceptual map which has two dimensions. The similarity between two factors more stronger, the configuration of two factors more close in the map. As a result of graph can show factors more clearly and let researchers understand easily. Therefore, Through Perceptual map could show the hidden structure or spatial relation between the factors, and achieve the classification result through the spatial difference.

3 Hybrid Decision-Making Support Model

This study proposes a Hybrid Decision-Making Support Model (HDMSM). As shown in Figure 2, the decision-making procedures according to HDMSM includes total four stages, namely, Selection, Relation, Evaluation and Decision, which are described below:

Selection. From literature, proper evaluation criteria and alternatives are selected for the goal in the decision making.

Relation. To understand the relation among different evaluation criteria, it is necessary to further use the DEMATEL method to analyze the degree of mutual influence among different criteria.

Evaluation. Based on the relation among different criteria as found in the stage II, a networking structure of evaluation is plotted. Then, according to the networking structure of evaluation, an ANP expert questionnaire is designed and distributed. Further, using ANP to analyze and calculate the weights and the priority ranking of the evaluation criteria.

Decision. Use ANP to process all the evaluation criteria and the alternatives, so as to pick out the most optimal alternative and calculate the Euclidean distance among the alternatives. Then, use MDS to analyze and find out the similarity and dissimilarity among all the alternatives.

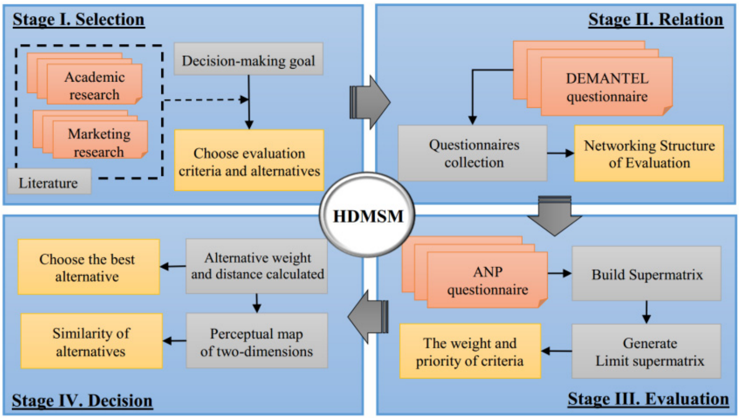


Fig. 1. Hybrid Decision-Making Support Model (HDMSM)

4 The HDMSM with Sample Demonstration

4.1 Criteria and Alternatives Selection

Before a multi-criteria decision making, it is necessary to select the criteria suitable for use as evaluation factors and to select the alternatives. The criteria can be obtained from past similar decision problems or related literature. In this study, we postulate five evaluation criteria (C1~C5) and five alternatives (A1~A5) for explaining the subsequent sample decision-making flow according to our HDMSM.

4.2 Establishment of Criteria's Relationship

After the criteria are selected, first distribute the DEMATEL expert questionnaire, so as to determine the relation and the degree of mutual influence among the criteria. The adopted scaling is the 0~3 scaling designed by Fontela & Gabus (1976); where, 0 indicates there is not any relation among the criteria, and 3 indicates there is significant relation among the criteria. Table 1 shows a sample questionnaire of DEMATEL.

Table 1. The Questionnire Sample of DEMATEL

	C1	C2	C3	C4	C5
C1	-	3	0	1	1
C2	0	-	2	2	0
C3	0	1	-	0	3
C4	2	0	0	-	2
C5	0	0	1	3	-

4.3 Plotting of Networking Structure of Evaluation

After the process using DEMATEL method, the criteria's relationship presented in the obtained results is used to plot a networking structure of evaluation, a sample of which is shown in Figure 3.

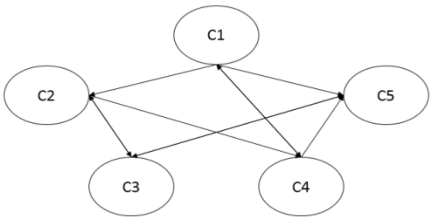


Fig. 2. The Sample of Networking Structure of Evaluation

Further, distribute the ANP expert questionnaire that is designed based on the networking structure of evaluation. For example, when use C1 as a basis of evaluation, a pairwise comparison is conducted among C2, C4 and C5. The questionnaire is rated on the Pairwise Comparison Scale proposed by Saaty (Saaty, 1980), as shown in Table 2.

Table 2. The Sample of Pairwise Comparison

Important <------------------------ Equal ------------------------> Important										
Scale	9:1	7:1	5:1	3:1	1:1	1:3	1:5	1:7	1:9	Scale
C2						V				C4
C2				V						C5
C4		V								C5

4.4 Criteria Weighting and Priority Ranking

The collected ANP expert questionnaires are calculated to acquire the eigenvectors of criteria and to form a Supermatrix. Through normalization of the Supermatrix and complex matrix multiplication, a Limit Supermatrix showing weights of the evaluation criteria can be obtained. A sample Limit Supermatrix is shown in Table 3.

Table 3. The Sample of Limit Supermatrix

	C1	C2	C3	C4	C5	Weight	Rank
C1	0.230	0.230	0.230	0.230	0.230	0.230	3
C2	0.338	0.338	0.338	0.338	0.338	0.338	1
C3	0.249	0.249	0.249	0.249	0.249	0.249	2
C4	0.033	0.033	0.033	0.033	0.033	0.033	5
C5	0.150	0.150	0.150	0.150	0.150	0.150	4
Total	1.000	1.000	1.000	1.000	1.000	1.000	-

4.5 Multidimensional Scaling and Alternative Selection

Normally, there is more than one alternative for a decision problem, for each alternative, different evaluation criteria usually have different importance levels. Therefore, after obtaining the criteria's priority ranking, it is necessary to further calculate the relative importance level of each alternative based on the evaluation criteria, so as to facilitate the subsequent alternative similarity analysis. Again, the pairwise comparison scale proposed by Saaty is used as the rating scale. Finally, the total weight and the priority ranking of each of the alternatives based on all criteria are obtained, as shown in Table 4.

Table 4. The Sample of Alternatives Priority

	C1	C2	C3	C4	C5	權重總合	Rank
A1.	0.473	0.170	0.111	0.134	0.554	1.442	1
A2.	0.059	0.055	0.423	0.095	0.102	0.733	4
A3.	0.036	0.117	0.162	0.043	0.231	0.589	5
A4.	0.149	0.396	0.271	0.480	0.064	1.361	2
A5.	0.283	0.262	0.033	0.249	0.048	0.875	3
	1.0000	1.0000	1.0000	1.0000	1.0000	-	-

Based on Table 4, the Euclidean distance between any two alternatives can be further calculated to create a Euclidean distance matrix, as shown in Table 5.

Table 5. The Sample of Euclidean Distance Matrix

	A1	A2	A3	A4	A5
A1	0.000				
A2	0.699	0.000			
A3	0.557	0.302	0.000		
A4	0.736	0.546	0.567	0.000	
A5	0.566	0.522	0.419	0.383	0.000

From the Euclidean distance matrix, we can find the coordinate positions of the alternatives in a second dimension and plot a perceptual map, as shown in Figure 4.

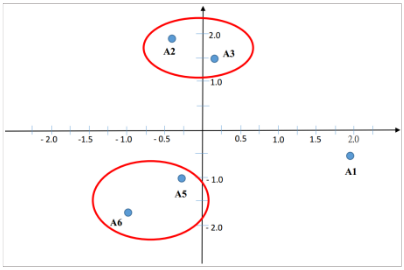


Fig. 3. The Sample of Perceptual Map

From the perceptual map, the decision maker can have a clear idea about the similarity and the dissimilarity among all the alternatives. Finally, along with the priority of the criteria and the alternatives to select the most optimal alternative.

5 Conclusions and Further Work

MCDM problem has always been a topic that can not be ignored in many different fields, such as management, social science and even engineering. The main objective of decision analysis is to help decision makers determine the evaluation criteria and find out the most optimal alternative for their decision problems.

This study combines three different MCDM methods, namely, DEMATEL, ANP and MDS, to propose a four-stage hybrid decision-making support model (HDMSM). This model can be used to effectively analyze the relational level among different evaluation criteria in a decision problem, and to find out the criteria that have most significant influence on the decision results. Then, the criteria and the decision alternatives are cross-analyzed to help the decision makers pick out the best alternative for the decision problem for execution.

The HDMSM can be applied to analyze decision problems of various issues, such as system introducing or business process reengineering. In future, HDMSM could be combined with other decision methods in order to improve the accuracy and effectiveness of this decision-making support model in handling decision problems.

References

1. Belton, V.: Multiple Criteria Decision Analysis Practically the Only Way to Choose. Strathclyde Business School (1990)
2. Huang, J.J., Tzeng, G.H., Ong, C.S.: Multidimensional Data in Multidimensional Scaling Using the Analytic Network Process. Pattern Recognition Letters 26(6), 755–767 (2005)
3. Saaty, T.L.: Analytic Hierarchy Process, Encyclopedia of Biostatistics. John Wiley & Sons, Inc. (1980)
4. Torgerson, W.S.: Multidimensional Scaling: I. Theory and Method. Pstchometrika 17(4), 401–419 (1952)
5. Yoon, K., Hwang, C.L.: Manufacturing Plant Location Analysis by Multiple Attribute Decision Making: Part I-Single-Plant Strategy. International Journal of Production Research 23(2), 345–359 (1985)
6. Simon, H.A.: The New Science of Management Decision. Englewood Cliffs (1977)
7. Opricovic, S., Tzeng, G.H.: Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS. European Journal of Operational Research 156(2), 445–455 (2004)
8. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Application. Springer (1981)
9. Buede, D.M., Maxwell, D.T.: Rank Disagreement: A Comparison of Multi-criteria Methodologies. Journal of Multi-Criteria Decision Analysis 4(1), 1–21 (1995)
10. Saaty, T.L.: Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in Complex World. RWS Publications (1988)
11. Keeney, R.L.: Value Focused Thinking. Harvard University Press (1992)
12. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publications (1996)
13. Ehrgott, M., Klamroth, K., Schwehm, C.: An MCDM Approach to Portfolio Optimization. European Journal of Operational Research 155(3), 752–770 (2004)
14. Shyur, H.J., Shih, H.S.: A Hybrid MCDM Model for Strategic Vendor Selection. Mathematical and Computer Modelling 44(7), 749–761 (2006)
15. Büyüközkan, G., Çifçi, G.: A Novel Hybrid MCDM Approach Based on Fuzzy DEMATEL, Fuzzy ANP and Fuzzy TOPSIS to Evaluate Green Suppliers. Expert Systems with Applications 39(3), 3000–3011 (2012)
16. Yang, Y.P., Shieh, H.M., Leu, J.D., Tzeng, G.H.: A Novel Hybrid MCDM Model Combined with DEMATEL and ANP with Applications. International Journal of Operations Research 5(3), 160–168 (2008)
17. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publisher (2001)
18. Wu, W.: Choosing Knowledge Management Strategies by Using a Combined ANP and DEMATEL Approach. Expert Systems with Applications 35(3), 828–835 (2008)
19. Yang, J.L., Tzeng, G.H.: An Integrated MCDM Technique Combined with DEMATEL for a Novel Cluster-Weighted with ANP Method 38(3), 1417–1424 (2011)
20. Fontela, E., Gabus, A.: Current Perceptions of the World Problematique. World Modeling-A Dialogue (1976)

A Study of Architecture and Protocol Designs for NREN Federation and Optimal Path

Hyunhun Cho, Jinhyung Park, Wonhyuk Lee, and Seung Hae Kim*

Dept. of KREONET Services, KREONET Center, KISTI, Daejeon, Korea
{hhcho,ntoskr,livezone,shkim}@kisti.re.kr

Abstract. According to analysis on recent changes in network users' values and their usage patterns, simple web information and one-way data transfer have declined. Instead, there have been rise in the use of multimedia, rising demand for security and personalization and increase in needs for liberal mobility. It is forecasted that thanks to these changes, current individual services would be gradually integrated in a converged form, and individual networks would evolve into an integrated intelligent network which meets users' diverse needs. As the Optical Transport Networks become more important, and network traffic is concentrated in the communication network components, it's been urgent to secure the survivability of the OTN. Therefore, it is critical to provide connectivity that end users want to the network service provider through optimal path allocation between the networks. As federation among international research networks becomes important, and there have been attempts to share a great volume of network resources in optimum conditions these days, it's been essential to select optimal path and share resources through information sharing among multi-domains. Considering these diversities, therefore, this study has tried to examine architecture and protocol to select optimal path among multi-domains.

Keywords: NREN, GFO(Global Federation Organizer), ICE(Information Control Element).

1 Introduction

As the OTN becomes more important due to increase in international collaborative research activities and supply of diverse services, and network traffic is concentrated in some communication network components, it's been urgent to secure the survivability of the OTN as well. As a result, communication network control technology has rapidly improved. Recently, in addition, there have been attempts to share a great volume of network resources in optimum conditions with a goal of federation among international research networks.

This study has considered approachable parts in advance from the perspective of the National Research & Education Network(NREN) except for limited factors which would be considered in conventional routing protocol under the multi-domain environment among NRNs around the globe. In addition, it has attempted to propose

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_39, © Springer-Verlag Berlin Heidelberg 2014

Global Federation Organizer(GFO) protocol which uses the Information Control Element(ICE) to support traffic engineering and plays a role of communication protocol with the GFO that can integrate and operate the ICE.

2 Path Computation Element(PCE)

In this section, the PCE in the Generalized Multi-protocol Label Switching(GMPLS) targeted to support multi-domain networks is described. Here, the PCE refers to an object to which constraints have been applied for the calculation of the network topology-based network path or route.

One of the inter-domain path calculation methods is to use the PCE. In this case, the PCE may not have a view on the complete network map, or it would be able to calculate loose routes from ingress to egress nodes. The calculated loose route is handed over to the Explicit Route Object(ERO) by the PCE. Using signaling, the loose routes are headed to the Link State Packet (LSP) which sets up requests along the path. The PCE can calculate paths based on the network diagram and Traffic Engineering Database(TED). The TED is formed by starting the IGP which is operated by TE extensions such as OSPF-TE and ISIS-TE, and it would include additional information including LSP path and traffic statistics. Because the PCE can calculate path in different types, it can provide traffic engineering functions in a GMPLS or MPLS network.

There are several ways to use the PCEs in single domain, single layer, multi-domains and multi-layers, but the PCE can be configured considering all layers in single domain and dynamically manageable domain with a goal of finding a method to use network resources efficiently. The reason why the PCE is used in this kind of situation is to discover a way to use network resources in an efficient manner.

3 GFO and ICE Architecture

Global Federation Organizer(GFO) and Information Control Element(ICE) can be independent objects in the independent module on the ingress node or a certain network server. The major role of the GFO and ICE is to handle the path calculation request from the ingress node or router.

3.1 GFO Architecture

- ICE Information Repository (IIR): Delivers information to the ICE in each domain at the storage of basic domain information and path calculation and requests the calculation of all paths
- ICE Management System (IMS): Registers, deletes and revises ICE information in the IIR
- Path Calculation Processor (PCP): Calculates optimal paths

3.2 ICE Architecture

- Route Computation Client(RCC): Edge/ingress node
- Route Computation Element(RCE): Calculates all paths based on information received from the IIR
- Traffic Engineering Database(TED): Stores network domain resources and topology

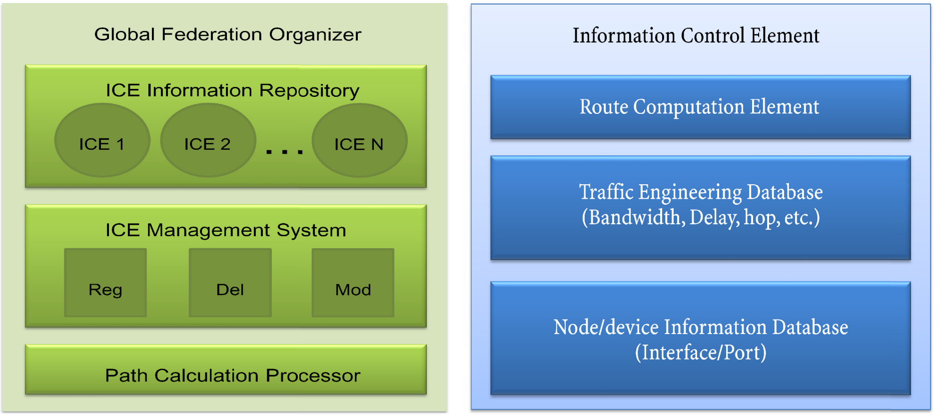


Fig. 1. Global Federation Organizer & Information Control Element Architecture

4 GFO Protocol

4.1 Protocol Architecture

In order for the ICE which exists in each domain and the GFO which controls the ICEs to deliver information, communication protocol is required. The GFP is request/response-type protocol which is operated based on the TCP. Therefore, it supports communication reliability and in-order delivery and uses the functions provided by the TCP as well to handle all security-related problems. In addition, the ICEP is session-based protocol which is similar to the Session Initiation Protocol (SIP). In other words, the ICE and PCC open session, negotiate parameters and learns assistance capabilities through communication. All messages exchanged between the GFO and ICE are traded within this kind of session, and the exchanged messages are as follows:

Table 1. Types of GFP Message

GFP Messages	Description
Open	Initiates and maintains the ICE and GFP sessions
Keepalive	
GFO_RReq	A message forwarded from the GFO to the ICE to request path calculation in each domain / includes destination information as well
GFO_RRep	A message forwarded from the ICE to the GFO as response to the path calculation request / sends negative response if more than one path which satisfies the request is included or if no satisfying path is available
GFO_Ntf	A message forwarded from the GFO to the ICE (or vice versa) / means that a certain event has taken place
GFO_Err	A message forwarded when a protocol error occurs
Close	A message used to close the GFP session

The GFO can have at least one ICE and GFP session each, and the GFP message is a transmission unit which should not be interleaved. For example, to close the session after forwarding the GFO_RReq message, the ICE must send the CLOSE message after the GFO receives all GFO_RRep messages.

4.2 GFP Message

Common header

Figure 2 represents the header format of the GFP message while the table below explains the field of the header in the GFP message.

Open

The GFP Open is a message to set the GFP-ICE session, and the first message to be transmitted must be the Open message.

Keep_alive

Keep_alive is a message used to keep the GFO or ICE session active. It is used for ACK which represents the normal reception of the Open message and acceptance of the GFP session. It is deemed that the GFP session is closed unless the Keep_alive or other GFP message is received for a certain period of time.

Route Computation Request(GFO_RReq)

It is a message forwarded from the GFO to the ICE. It sends final destination information to the ICE in the domain which includes at least one path calculation request and destination uses '3' for the Message-Type field value.

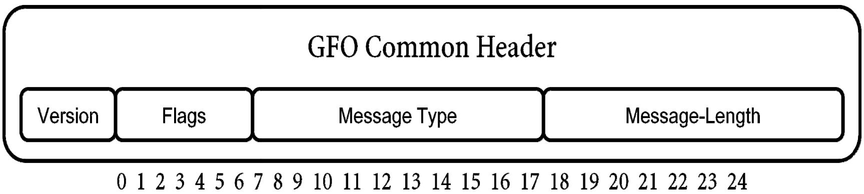


Fig. 2. Common Header of GFP Message

Table 2. Filed Contents of GFP Messages

<table>
<tr><th>Field</th><th>Contents</th></tr>
<tr><td>Version</td><td>Represents GFP version / expressed in 3 bits</td></tr>
<tr><td>Flags</td><td>Undefined, should be filled with ‘0’ / expressed in 5 bits</td></tr>
<tr><td>Message-Type</td><td>Defined in 8 bits as stated below:
<table>
<tr><th>No.</th><th>Meaning</th></tr>
<tr><td>1</td><td>Open</td></tr>
<tr><td>2</td><td>Keep_alive</td></tr>
<tr><td>3</td><td>Route Computation Request</td></tr>
<tr><td>4</td><td>Route Computation Reply</td></tr>
<tr><td>5</td><td>Notification</td></tr>
<tr><td>6</td><td>Error</td></tr>
<tr><td>7</td><td>Close</td></tr>
</table></td></tr>
<tr><td>Message-Length</td><td>The total length including the common header is variable/byte unit</td></tr>
</table>

Route Computation Reply(GFO_RRep)

Route Computation Reply is a message forwarded from the ICE to the GFO which requested path calculation. The ICEs which received the GFO_RReq message include the results of calculation on the path including destination if the final destination is in its domain or on all other paths including ingress/egress in the domain if the destination is not in its domain.

Notification(GFO_Ntf)

GFO_Ntf is a message used in the GFO or ICE to notify the occurrence of a certain event. It must have at least one notification object. It can include the RP object if the notification is related with a particular path calculation request

Error(GFO_Err)

GFO_Err is a message forwarded in the occurrence of protocol error or failure to meet the GFP standards such as the reception of the following messages; malformed message, message without a mandatory object, policy violation, unexpected message and unknown request reference.

Close Message

It is a message forwarded by the GFO or ICE. It is used to close the GFP session.

5 Performance Evaluation

5.1 Consist of Testbed

In this chapter, it has been applied to the GFO structure which has been proposed based on the multi-domain network, and its effectiveness has been investigated. To verify a model which can calculate optimal path based on the typical transfer characteristics, testbed has been organized as shown in Figure 3 below:

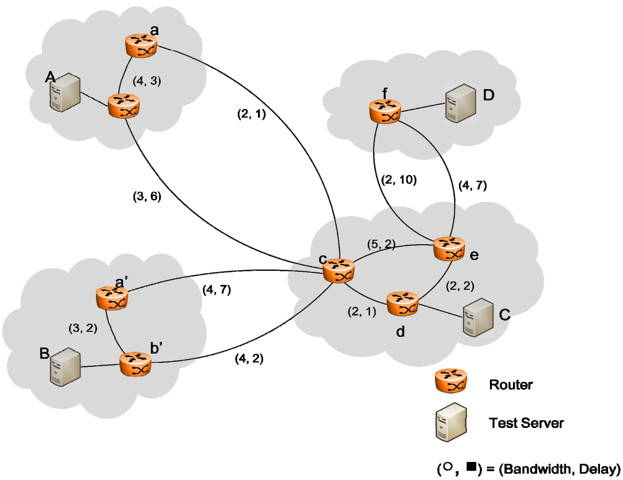


Fig. 3. Testbed of GFO

5.2 Configuration and PerformancTestbed

The After testbed was organized based on the NREN, a test was performed by selecting two cases among current active research fields; ‘transfer of large-volume data’ which is limited in transfer rate and band and ‘transfer of high-resolution(HD or higher) images’ which are sensitive to delay time. In terms of testbed configuration, four domains linked with 1Gbps and 10Gbps have been configured as shown in Figure 3 to create multi-domain environment. When the minimum hop-based path setting is compared to the proposed path setting model in the event of setting optimal path which takes test servers 'A' and 'B' as source and the test server 'D' linked as destination, they are compared and analyzed from the perspective of transfer rate and delay time (delay, jitter).

For data transfer in which transfer rates and bandwidth are more important than delay just like transfer of large-volume data, transfer from the source 'A' to the destination 'D' has been tested. For data transfer in which transfer delay such as real-time transfer of HD(or higher level) media is more important, transfer from the source 'B' to the destination 'D' has been analyzed in assumption that there would be no changes in network state such as dramatic increase of traffic.

When high-resolution images which would be used in teleconference or real-time education are transmitted, on the contrary, it is not easy to recover the data. Therefore, it is desirable to approach the transfer of video and audio data which are needed for real-time collaboration based on delay or jitter instead of minimum hop or transfer rate.

Approximately 250Mbps of transfer rate has been detected in forwarding 900MB data from the source 'A' to the destination 'D.' For comparison, it reveals a data transfer test on the path2 which was determined based on cost. Here, about 310Mbps of transfer rate was found. In other words, transfer rate improved by nearly 20%.

The results of the real-time transfer test on high-resolution media(HD or higher) are shown in Table 3, Test 2. Specifically, a data transfer test was performed on the path3 (source 'B' to the destination 'D'), and 3.2% of packet loss rate, 13ms of video jitter and 12ms of audio jitter were stated in the receiving area. In the delay-considered path4, packet loss rate was below 0.1m, video and audio jitters were 3ms each as shown in Table 3. The average of the test results are described in Table 3. When common data were transmitted, transfer rate improved by about 20% in path2, compared to path1. In case of real-time transfer of video data, packet loss decreased from 4,000 to 160 in path4, compared to path3. In jitter as well, audio and video jitters were 12ms and 3ms respectively, showing adequate network conditions for real-time data transfer.

Table 3. Comparison between Test Result

		Route	Throughput	Packet loss	Jitter (Video,Audio)
Test1	Path 1 (Min hops)	A-a-c-e-f-C	31.4MB/s		
	Path 2 (Throughput)	A-a-c-d-e-f-C	38.1MB/s		
Test2	Path 3 (Min hops)	B-a'-c-e-f-C	-	3.0%	13m/s, 12m/s
	Path 4 (Delay)	B-a'-c-d-e-f-C	-	0.0%	3m/s, 3m/s

6 Conclusion

At present, a network consists of individual management domains due to scalability, manageability and commercial concerns. In fact, because it is unable to get complete end-to-end path information from a particular user, it is extremely difficult to calculate optimal path up to the destination. In multi-domain environment, therefore, it is a good idea to collect node information for path calculation and have a hierarchical path node calculation structure to share the information. Since it is hard to search appropriate paths with all path information in a single node, it is important to find a border between the exit point and next domain in the domain as soon as possible. For this, architecture in which network information control node, ICE and top path calculation node exist is needed.

After designing the ICE and GFO-based structure efficiently in consideration of national research network environment and applying it to the multi-domain network, it was possible to use resources in an efficient manner, select optimal path among multi-domains and optimize end-to-end connection. In addition, a customized path selection model has been applied through testbed organization for logical multi-domain environment. Then, an optimal path calculation model targeted to provide fast network services are proposed by concentrating on the design of architecture for optimal path selection through minimum information transfer among nodes in multi-domain environment as well as for the efficient use of resources.

References

1. Bonerjee, A., Drake, J., Lang, J.P., Turner, B.: Generalized Multiprotocol Label Switching: an Overview of Routing and Management Enhancements. IEEE Commun. Magazine 39(1), 144–150 (2001)
2. Nishioka, I., Ishida, S., Iizawa, Y.: End-to-end path routing with PCEs in multi-domain GMPLS networks. In: 4th Int. Conference on IP & Optical Network, Tokyo, Japan (2008)
3. Bouabdallah, N., Pujolle, G., Dotaro, E., Le Sauze, N., Ciavaglia, L.: Distributed Aggregation in All-Optical Wavelength Routed Networks. In: 2004 IEEE International Conference on Communications, pp. 1806–1810 (June 2004)
4. Seok, Y., Lee, Y., Choi, Y., Kim, C.: Explicit Multicast Routing Algorithms for Constrained Traffic Engineering. In: Proceedings of the 7th ISCC, pp. 455–461 (July 2002)
5. Cha, M., et al.: Path Protection Routing with SRLG Constraints to Support IPTV in WDM Mesh Networks. In: 25th IEEE International Conferenceon Computer Communications, INFOCOM 2006, pp. 1–5 (2006)

Cognitive Diagnosis of Learning Path in Geography Based on Rule Space Model

Chang Wen-Chih* and Wu Shi-Wei

Dept. Information Management
Chung Hua University
Hsinchu, Taiwan
earnest@chu.edu.tw

Abstract. One of the most important advantages of E-Learning is in providing timely contents through Internet. In a similar manner, real-time assessment systems allow students to acquire instant feedback of their performance. In this article, in addition to informing students about their misconceptions on the subject that they are learning, the proposed diagnostic scheme assesses their learning behaviors and provides appropriate remedial suggestions. Therefore, the knowledge attributes tree-like structure and the learning paths are proposed and analyzed in this paper helps instructors to realize the learning progress of the entire class. Most diagnostic assessment approaches spend a relatively longer time by adjusting and inducing knowledge conceptions. The proposed Geography learning applied the rule-space model to help instructors and students to realize and diagnose learning status, attitude and remedial needs. An experiment using our system is performed to reveal the advantage of our computation method in diagnosing learning problems.

Keywords: Index Terms—Diagnostic Assessment, Rule-Space Model, e- Learning, Geography Learning.

Introduction

Owing to the popularity of e-learning, learning materials and social group communication platforms are now widely established on the Internet, including wireless networking environments. In education environment, the learning cycle includes both instruction and assessment, whether in an e-learning environment or a traditional classroom. Instruction makes students learning by the teachers and the learning material. Assessment realized the students learning ability and cognitive learning status. It is possible that an instructor misses some key conceptual points in the teaching of a course. Another possible is instructors did not put all the learning content in the assessment for the time and paper sheet limitation.

Cognitive development of learner is too abstract to measure directly and thus it is typically measured indirectly using test scores. However, test scores provide a poor

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_40, © Springer-Verlag Berlin Heidelberg 2014

indication of knowledge which is composed of both conceptions and misconceptions held by a student. For this reason, rule assessment approaches have been developed [1] to diagnose misconceptions hold by students and thus to provide valuable information for planning remediation.

Student indices [2], [3], [4], [5], [6], [7] apply rules to detect mistakes cause by abnormal behavior of students in tests. However, student indices in the Student-Problem Chart (i.e., SP Chart) can only find individual students, instead of what problems may occur [8]. The SP Chart cannot diagnose the misconceptions of students, nor provide effective remedial content for students.

Tatsuoka and her associates [6] have proposed a rule-space model and have successfully adopted it to diagnose misconceptions held by pupils in signed number and fraction arithmetic. However, the rule-space model involves complicated test items, and may take a long time to apply.

Related Works

In section 2, we introduced some related researches about diagnostic assessment methods. Tatsuoka [9] designed a rule space model for diagnosing the abstract sources of misconceptions among students. Many investigations have applied the rule-space model in different research. Menucha B. *et al.* [1] proposed a rule space model for cognitive analysis of math learning behavior among students, and employed their model to measure knowledge of Algebra among students. A 32-item test was performed on 231 eighth and ninth graders. The rule space model classifies examination candidates by their level of knowledge. Sheehan *et al.* [10] developed a modification to the diagnostic classification procedure of the rule space model. The underlying cognitive model of generalized problem-solving skills can be performed to determine the comprehensive set of knowledge states for examinee classification. Irvin R. Katz *et al.* [11] applied the rule space model in a semantically-rich domain (Architecture knowledge) with three attributes, namely (1) understanding, (2) solving, and (3) checking. Understanding comprises processes for building an initial representation of an item. Solving consists of processes, goals, and actions performed based on the goals. Checking involves determining whether goals have been attempted and satisfied.

Geography Based on Rule Space Model

Geography learning is one of the social science subjects. Teachers use figures, table and demonstration for students. However, geography learning is similar with other subjects; teachers are not familiar with realizing the knowledge status in the whole classes. In this research, we applied Rule Space Model in Geography learning. Rule Space Model is not easy to apply in a whole textbook, when the teachers have to spend a lot of time to design and plan the knowledge structure. The detail knowledge attributes are also not easy to define in a short time. This research applied Rule Space model in Geography learning in one of the chapters in central part Taiwan.

First, we try to design the learning content. The learning content is designed as tree structure for analyzing the knowledge states. The teacher put the basic knowledge attribute in the beginning and put the advanced knowledge attribute in the middle and end of the tree structure. Each knowledge attribute is named as A + number. A means attribute and number means the sequence order. Figure 1 shows the knowledge structure in Industry (II) of Geography. It composed of Industry Category, Characteristic, Taiwan's industrial distribution, Taiwan's industrial development stage, Business, International Business, Taiwan International Business and Taiwan Tourism.

The experiment applied social area Geography subject, we divided the content into 8 knowledge attributes. We called the first knowledge attribute as A1 for concept 1. The entire knowledge concepts are listed in TABLE 1.After the tree-like structure, we create the adjacency matrix from Figure 1. Table 2 shows the adjacency relations. “1” means the two knowledge attributes has relation. “0” means the two knowledge attributes has no relation.

Table 1. Knowledge attribute table

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	1	0	0	0	0	0	0
A2	0	0	1	1	0	0	0	0
A3	0	0	0	0	1	0	0	0
A4	0	0	0	0	1	0	0	0
A5	0	0	0	0	0	1	0	1
A6	0	0	0	0	0	0	1	0
A7	0	0	0	0	0	0	0	0
A8	0	0	0	0	0	0	0	0

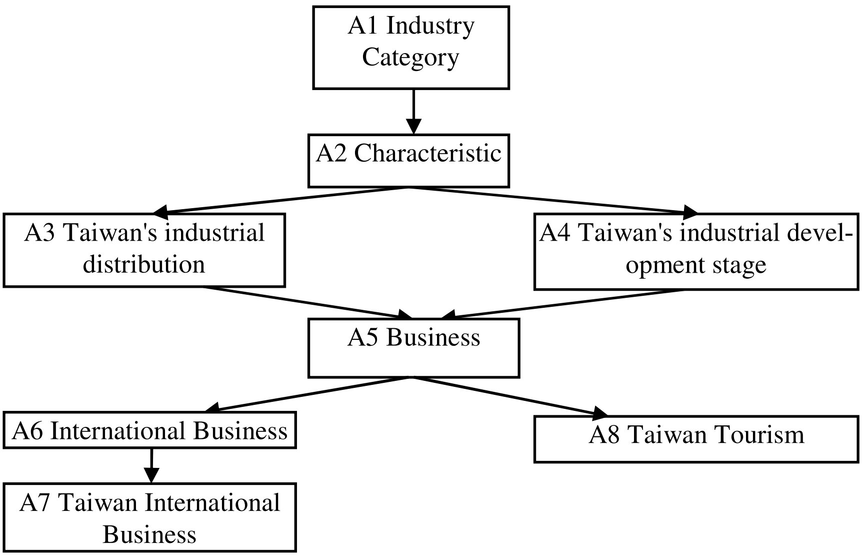


Fig. 1. Knowledge attributes tree structure

We applied Rule Space Model to find out the cognitive concept design in the junior high school students Geography learning in Taiwan. We designed the experimental group and control group in this study. We designed the cognitive concept and learning map, which can be used to find out the students learning path.

Table 2. Adjacency Matrix

No.	Knowledge Attribute	Content
A1	Industry Category	Three kinds of industries
A2	Characteristic	raw materials, power, labor, marketing, transportation
A3	Taiwan's industrial development stage	Taiwan's industrial development Characteristics in different decade
A4	Taiwan's industrial distribution	The distribution and presentation of regional specialties Taiwan's industrial zone
A5	Business	the business activity's classification
A6	International Business	The basic conditions for international trade and concepts
A7	Taiwan International Business	Current status of Taiwan in international trade
A8	Taiwan Tourism	Current status of Taiwan's tourism industry

Before learning activities, we take a pretest to experimental group (O_1) and control group (O_2). Then they have the same teaching(∆). However, the experimental group students have individual learning advice when they completed the pretest. The experimental group teacher will explain how to read and use it.

After the teaching activity, the experimental group and control group have posttest（O_3、O_4）.In order to collect and analysis their opinion, we have interview to some experimental group.

Geography Learning Experiment Analysis

The experimental group is composed of 34 students. There are 20 male and 14 female. The control group has 33 students. There are 19 male and 14 female. After the analysis, the experimental group and control group have significance. Table 3 shows the posttest T-test result.

Table 3. Posttest Average and T-test

	Students	Average	Standard deviation	T	Significance
Experimental Group	34	64.84	25.00	2.396	0.019*
Control Group	33	49.91	25.31		

*P<0.05.

Table 4. The knowledge status distribution (N=34)

Knowledge State	Description	Number	Percentage
KS1	All the knowledge are not mastered	4	12%
KS2	Learners is mastered A1	28	82%
KS3	Learners is mastered A1、A2	20	59%
KS4	Learners is mastered A1、A2、A3	16	47%
KS5	Learners is mastered A1、A2、A4	17	50%
KS6	Learners is mastered A1、A2、A3、A4	14	41%
KS7	Learners is mastered A1、A2、A3、A4、A5	11	32%
KS8	Learners is mastered A1、A2、A3、A4、A5、A6	7	21%
KS9	Learners is mastered A1、A2、A3、A4、A5、A6、A7	7	21%
KS10	Learners is mastered A1、A2、A3、A4、A5、A8	9	26%

The advanced analysis for the experimental group is collected in Table 4. We defined 10 knowledge states named KS1 to KS10. Each knowledge state indicates specific knowledge attributes. For example, KS1 means "All the knowledge are not mastered." and KS4 means "Learners is mastered A1、A2、A3."Combined with the experimental group data, Table 8 shows the percentage of each knowledge status.

According to result, there are four possible learning path is the following.

Path 1：KS 1 → KS 2 → KS 3 → KS 4 → KS 6 → KS 7 → KS 8 → KS 9
Path 2：KS 1 → KS 2 → KS 3 → KS 5 → KS 6 → KS 7 → KS 8 → KS 9
Path 3：KS 1 → KS 2 → KS 3 → KS 4 → KS 6 → KS 7 → KS10
Path 4：KS 1 → KS 2 → KS 3 → KS 5 → KS 6 → KS 7 → KS10

Conclusion

In the traditional classroom and e-Learning classes, students need an effective diagnostic assessment system for providing useful remedial information, supporting learning status of the whole class. The proposed Rule Space Model applied in Geography learning. The proposed Geography tree-like structure provides efficient diagnostic assessment functions for students.

Acknowledgment. We would like to thank National Science Council and Chung Hua University. This research was supported in part by a grant from NSC 102-2511-S-216-001, Taiwan, Republic of China.

References

1. Menucha, B., Anthony, K.E., Tatsuoka, K.K.: Diagnosing Knowledge States in Algebra Using the Rule Space Model. Educational Testing Service Report (1992)
2. Drasgow, F.: Choice of test model for appropriateness measurement. Applied Psychological Measurement 6(3), 297–308 (1982)
3. Harnisch, D.L., Linn, R.L.: Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement 18(3), 39–87 (1981)
4. Levine, M.V., Rubin, D.B.: Measuring the appropriateness of multiple-choice test scores. Journal of Educational Statistics 4(4), 269–290 (1979)
5. Sato, T.: The construction and interpretation of S-P tables. Meiji Tosho, Tokyo (1975) (in Japanese)
6. Tatsuoka, K.K., Tatsuoka, M.M.: Detection of aberrant response patterns. Journal of Educational Statistics 7(3), 215–231 (1982)
7. Wright, B.D., Stone, M.H.: Best test design, Rasch measurement. MESA Press, Chicago (1979)
8. Sato, T.: Introduction to S-P Curve Theory Analysis and Evaluation. Meiji Tosho, Tokyo (1985) (in Japanese)
9. Tatsuoka, K.K.: Rule Space: An Approach for Dealing with Misconceptions based on Item Response Theory. Journal of Educational Measurement 20(4), 345–354 (1983)
10. Sheehan, K.M., Tatsuoka, K.K., Lewis, C.: A Diagnostic Classification Model for Document Processing Skills. Educational Testing Service Report (1993)
11. Katz, I.R., Martinez, M.E., Sheehan, K.M., Tatsuoka, K.K.: Extending the Rule Space Model to a Semantically-Rich Domain: Diagnostic Assessment in Architecture. Educational Testing Service Report (1993)

Secure Resource Synchronization of Mobile Peer-to-Peer Techniques: Experiments on the Android Platform

Chih-Kun Ke and Zheng-Hua Lin

Department of Information Management,
National Taichung University of Science and Technology,
40401 Taichung, Taiwan, R.O.C.
{ckk,s1800b102}@nutc.edu.tw

Abstract. Traditional resource synchronization processes exchange files between PCs and enterprise mainframes. In recent years, smart devices have increased employee mobility, allowing them to carry work in a ubiquitous environment. Resource synchronization has thus become an important focus of investigation, particularly in terms of which network transmission protocol is most suitable for use by smart devices. A peer to peer (P2P) network is a type of decentralized and distributed network architecture. Individual peers in the network act as both suppliers and consumers of resources. In a ubiquitous environment, a reasonable solution for file resource synchronization is to construct a mobile peer to peer (MP2P) network based on a distributed peer to peer architecture that provides a secure platform. In this study, we explore network transmission protocols, including Juxtapose (JXTA), Session Initiation Protocol (SIP), Bluetooth, and WiFi Direct, in order to construct a MP2P network of Android-based smart devices for file resource synchronization. The experiments demonstrate that the JXTA protocol provides more functionality for Android-based smart devices.

Keywords: Ubiquitous, mobile peer to peer (MP2P), resource synchronization, Juxtapose (JXTA), Android.

1 Introduction

Enterprise resources planning (ERP) systems that integrate smart devices have improved enterprise mobility, so when employees execute specific business processes, they can access the required file resources through smart devices from almost anywhere. Therefore, ERP systems combined with smart devices in order to carry out business processes have gradually become an important trend. However, a critical concern is how to effectively and securely share information over a secure platform using smart devices as an auxiliary tool. In addition, satisfying the comprehensive requirements of relevant employees is another issue worth exploring. A viable option in this case would be to construct a Mobile Peer to Peer (MP2P) sharing environment based on the Distributed Peer to Peer (DP2P) architecture over Cloud Computing that provides a secure platform for file resource sharing. Researchers have used peer to

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_41, © Springer-Verlag Berlin Heidelberg 2014

peer protocols to implement applications in several domains. Rajkumar et al. (2012) [1] used JXTA to construct a mobile peer to peer healthcare management system allowing patient and healthcare workers to communicate messages and transmit data. Tsai et al. (2009)[2] used JXTA to design a mobile peer to peer social network application which allows students to transmit file resources and share relevant course information with mobile devices. Barolli et al. (2011) [3] used JXTA to construct a ubiquitous peer to peer platform for resource sharing and collaborative processing. The platform communicates with remote devices by JXTA, and controls a learning environment for more effective learning. Lin et al. (2011, 2012) [4][5] used SIP to construct a mobile learning environment which allows students and teachers to learn at any time. Cao et al. (2012) [6] used a Bluetooth and 3G mobile network to design a blood glucose monitoring system which automatically detects the blood glucose status of diabetics. The status is then immediately transmitted to a healthcare center by Bluetooth transmission protocols. If a patient's blood glucose status is abnormal, the system uses GPS to obtain their location, and sends an alert to the healthcare center. Huang et al. (2011)[7] composed the IEEE1588 precision timing protocol, WiFi Direct, and used a time of arrival measurement of distance algorithm to predict the distance of a wireless real-time location.

In this study, we implement an MP2P resource sharing mechanism to share files over a secure platform in a ubiquitous environment. Juxtapose (JXTA) could then be used to transfer the shared file resources. When employees execute specific business processes, they can use the MP2P features to synchronize the system file resources through a smart device from anywhere, and thus improve the efficiency of file resources sharing. During file resource syncing, a log file of resources is updated in order to maintain the consistency of the file resources. All file resources' encryption/decryption follows the Advanced Encryption Standard (AES) of symmetric data encryption [8]. The contribution of this study is to offer a resource sharing mechanism to enhance the effectiveness of enterprise systems. In addition, enterprise mobility will be improved, allowing employees to access file resources through smart devices from anywhere.

The remainder of this paper is organized as follows. Section 2 introduces the experimental resource synchronization application for Android-based smart devices and relevant discussions. Finally, in Section 3, we present our conclusions and indicate the direction of future work.

2 Experimental Resource Synchronization Application for Android-Based Smart Devices

This section describes the prototype experimental applications used on Android-based smart devices. Using the JXTA protocol, we implement the application providing resource synchronization among multiple smart devices, including one device to one device, and one device to multiple devices. However, the Bluetooth and WiFi Direct protocol Android APIs do not support one device to multiple device resource synchronization. Therefore, we implement the applications which only provide the function of one device to one device resource synchronization with Bluetooth and WiFi Direct protocols. In addition, because the SIP protocol does not support file resource sharing, we do not implement the experimental SIP application for evaluation.

2.1 Operational Steps of the Application Based on the JXTA Protocol

The JXTA-based application builds an experimental environment and executes the operations to construct a JXTA peer, and includes seven steps [2], as shown in Figure 2.

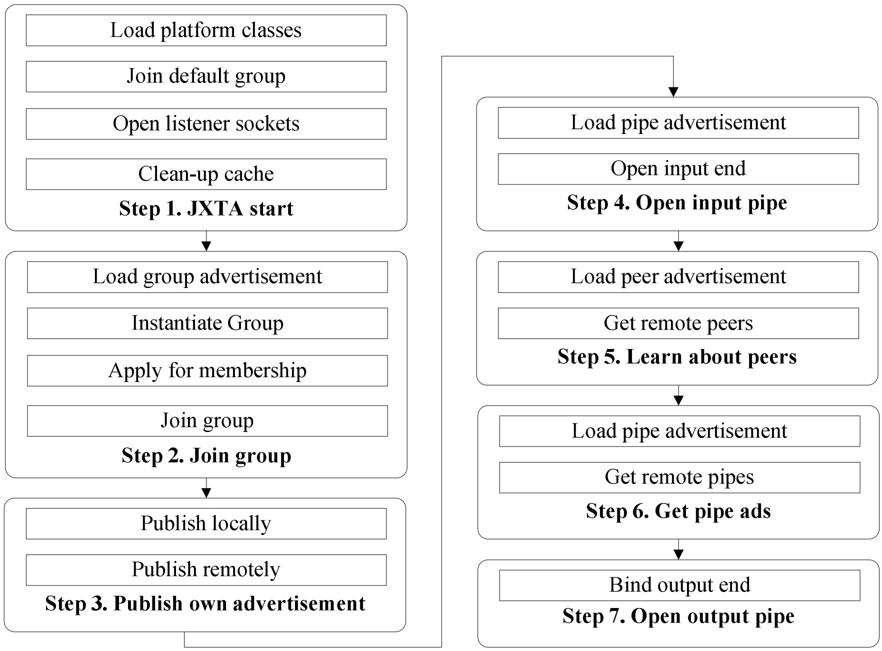


Fig. 1. JXTA network construction [2]

Start JXTA step loads the JXTA class library and constructs information for connection, including TCP port, broadcasting, peer configuration description, the IP address of Rendezvous Peer, peer type, etc. Based on this information, the application constructs a peer, which is named by user account, and starts to listen to the socket. After opening the JXTA platform, the peer loads a peer group advertisement when joining a peer group. If a peer wants to join a peer group, but there are no existing peer groups, the peer uses class library to construct a peer group advertisement. The user can also modify the content of the advertisement. If a peer group exists, the peer will load the peer group advertisement. A peer will send a self-advertisement to the rendezvous peer in a peer group. The rendezvous peer is usually the owner of a peer group, and manages that peer group. The advertisement of a peer lists the service or file resource provided by the peer. Announcing this advertisement serves to notify other peers that it provides that service or file resource. Each peer must query the file resource parameters in an advertisement before accessing the file resources. Open input pipe step constructs the transmission channel of the advertisements and listens to determine if the channel receives the advertisement or not. This channel can only transmit the advertisement message. Learn about peers step reads the file resource table provided by the Rendezvous Peer. The file resource table lists file resources shared by each peer in a peer group. Before constructing the communication channel with other peers, a peer must search for and obtain other peers' advertisements in order to configure the channel. If a peer wants to transmit a message, it can obtain other peers' message receiving channel from the channel advertisement. It uses the pipe binding protocol to construct the communication channel, and sends the message to the destination peer. Figure 2 shows the resource synchronization between three smart devices.

Fig. 2. Resource synchronization between three smart devices

2.2 Relevant Discussions of Experiments

Gao et al. (2012) [9] proposed a collaborative middleware application in a Bluetooth communication network. However, there has are still some weak points, with Bluetooth properties limiting the scope of data transmission In Bluetooth version 2.0, the scope of data transmission is from 1 meter to 30 meters, and the data transmission rate is from 1 Mbps to 3 Mbps. These properties are suitable for a small scope environment and longer delay times in data transmission. Although Bluetooth version 3.0 improves the data transmission rate up to 24 Mbps, it still influences the data transmission rate in a collaborative environment. WiFi Direct, on the other hand, has a high speed data transmission rate and a larger data transmission scope. Huang et al. (2011) [7] measured the data transmission delay times of WiFi Direct devices without a wireless access point. Using WiFi Direct not only transmits data at higher speed, but also transmits data among WiFi devices more easily. Its maximum data transmission rate is 250 Mbps, and the maximum transmission distance is 200 meters. WiFi Direct is better than Bluetooth in terms of both data transmission rate and scope. However, the limitation of WiFi Direct is that the data transmission can only be enforced in a local network environment constructed by WiFi devices. In addition, when devices transmit data by WiFi Direct, they can either receive or send data separately, and cannot enforce two operations at the same time. Therefore, WiFi Direct may not be suitable for use in a collaborative sharing environment. Lin et al. (2011) [4] designed and implemented an SIP-based distance education system. However, SIP does not offer a file sharing function. If users want to share files in this education system, third-party software is needed in order to add a file sharing function. In terms of functionality, JXTA does not support multimedia communication, but does provide a complete collaborative resource sharing function. Barolli et al. (2011) [3] deployed JXTA in a peer to peer platform for file resource sharing between heterogeneous devices. The peer group mechanism strengthens the security of data exchange and collaborative processing. The advertisement of a peer group shows service and file resource sharing in the platform. A peer in the peer group can query the desired service or file resource index. The peer can then access the support service or file resource. In this study, the experiments demonstrate that the JXTA protocol provides more functionality to Android-based smart devices, including one device to one device, and one device to multiple devices transmission. The WiFi direct protocol facilitates high speed resource synchronization, and the Bluetooth protocol facilitates small file (1KB to 100KB) with high speed resource synchronization, but when the file size is larger, the transmission performance becomes poor. In addition, WiFi direct and Bluetooth protocols can only be enforced in one device to one device transmission on Android-based smart devices.

3 Conclusion

In this study, we implement a resource sharing mechanism using MP2P to sync files over a secure platform in a ubiquitous environment. When employees execute specific business processes, they can use the MP2P features to synchronize system file resources through a smart device from any location, and thus improve the efficiency of file resource sharing. When the file resources are synced, a resources log file is updated in order to maintain the consistency of file resources. All file resources' encryption/decryption follows the AES of symmetric data encryption. The contribution of this study is to provide a resource sharing mechanism to enhance the effective and secure sharing of files in an enterprise system. In addition, enterprise mobility is improved, allowing employees to access file resources through smart devices from any location. The experiments demonstrate that the JXTA protocol provides more functionality to Android-based smart devices. Future works focus on increasing the efficiency of the shared file resources by assessing the performance of each smart device before the syncing processing. In addition, this work will add collaboration analysis methods to the proposed mechanism in order to make it more intelligent and to improve its effectiveness.

Acknowledgments. This research was supported in part by the Industrial Technology Research Institute and the National Science Council of Taiwan (Republic of China) with a NSC grant 102-2410-H-025-017.

References

1. Rajkumar, R., Iyengar, N.C.S.N.: JXTA: A Technology Facilitating Mobile P2P Health Management System. Osong Public Health and Research Perspectives 3(3), 165–169 (2012)
2. Tsai, F.S., Han, W.C., Xu, J.W., Chua, H.C.: Design and development of a mobile peer-to-peer social networking application. Expert Systems with Applications 36(8), 11077–11087 (2009)
3. Barolli, L., Xhafa, F.: JXTA-Overlay: A P2P Platform for Distributed, Collaborative, and Ubiquitous Computing. IEEE Transactions on Industrial Electronics 58(6), 2163–2172 (2011)
4. Lin, H.Y., Wang, Y.J.: A Design and Implementation of SIP-Based Distance Education System. Procedia Engineering 15, 2123–2127 (2011)
5. Lin, H.Y., Wang, Y.J., Su, Z.Y.: Implementation on the Mobility of SIP-based M-leaning. Procedia Engineering 29, 3836–3840 (2012)
6. Cao, D., Cai, Y.Y., He, X.H., Wang, Q.X.: Continuous Glucose Monitoring System Based on Smart Phone. Procedia Engineering 29, 3894–3898 (2012)
7. Huang, J.S., Liu, X.X., Chen, Z.J.: The Human Positioning System Based on the WiFi Direct and Precision Time Protocol. In: Proceedings of the International Conference on Transportation, Mechanical, and Electrical Engineering, China, pp. 1580–1584 (2011)
8. Alanazi, H.O., Zaidan, B.B., Zaidan, A.A., Jalab, H.A., Shabbir, M., Al-Nabhani, Y.: New Comparative Study Between DES, 3DES and AES within Nine Factors. Journal of Computing 2(3), 152–157 (2010)
9. Halepovic, E., Deters, R.: The Costs of Using JXTA. In: Proceedings of the Third International Conference on Peer-to-Peer Computing, Sweden, pp. 160–167 (2003)
10. Gao, F., Liu, Q., Hope, M., Fan, T.R.: Improving communications with Bluetooth enabled collaborative middleware in MANET. The Journal of China Universities of Posts and Telecommunications 19(1), 100–107

A Search of Topics in Medical Journals: Applying Information Retrieval Approach to European Journal of Radiology

Deng-Yiv Chiu [1,*], Agus Andria[1], C.J. Juan[2], and H.C. Chiu[2]

[1] Department of Information Management, Chung Hua University, Taiwan, ROC
{chiuden,d09803024}@chu.edu.tw
[2] Diagnostic Radiology Ultrasound Department,
National Defense Medical School and Hospital, Taiwan, ROC
{peterjuancj,zinniachiiou}@yahoo.com.tw

Abstract. We apply information retrieval to search topics in European Journal of Radiology. Genetic algorithm is used for optimization. We also explore related features to each topic and measure the relationship between topic and its related features. The result shows that the topic search constructed with the aid of genetic algorithm featured a certain degree of representativeness.

Keywords: Topic search, Information retrieval, Genetic algorithm, Independent chi-square.

1 Introduction

Finding the relevant or useful information from the information storage is an important and typical task. There are many research dedicated to retrieve the relevant topic or useful information from the information resources. List information can be used to infer topics of expertise for individual twitter users [1]. A hybrid heuristic scheme has been presented for re-ranking academic papers retrieved from standard digital libraries such as the ACM Portal [2]. Topic map also has been developed to visualize search result. [3].

Several bibliometric studies in radiology focus on trend of topic [4-6]. Instead for clinical purpose, radiology is useful in forensic duty. A method also has been proposed to perform full text searches of radiology reports [7]. It is practically inexpensive for large number of department.

In this work, we introduce a method to search topics in European Journal of Radiology by applying information retrieval. Genetic algorithm is also adopted to optimize parameters. Thus, we explore related features to each topic and measure the relationship between topic and its related features. Chi-square test is performed to examining independence of topics.

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_42, © Springer-Verlag Berlin Heidelberg 2014

2 Related Methods

2.1 Term Weighting, Vector Space Model and Similarity Computation

Term-weighting is to find out the importance of each term in the document. The *tf-idf* approach is often used for term weighting that considers the term frequency (*tf*) and inverse document frequency (*idf*). The importance of term (t_m) is proportional to the term frequency and inversely proportional to the document (d_i) frequency, which is expressed as follows:

$$w\left(t_m, d_i\right) = tf_{nm} * \log\left(N / df_m\right) \tag{1}$$

where tf_{nm} denotes occurrence count of term t_m in document d_n, N is the total count of documents, df_m is the count of documents containing term t_m.

The similarity between texts can be estimated when the texts are presented in a vector space model [8]. The cosine formula is adopted to determine the similarity between two objects expressed as follows. The similarity is higher if the value approaches 1, while the similarity is lower if the value approaches 0.

$$sim\left(\overline{o_i}, \overline{o_j}\right) = \cos\left(\overline{o_i}, \overline{o_j}\right) = \left(\overline{o_i} \cdot \overline{o_j}\right) \Big/ \left(\left|\overline{o_i}\right| \times \left|\overline{o_j}\right|\right) \tag{2}$$

where $\overline{o_i}$ is the vector of object o_{i}, $\left|\overline{o_i}\right|$ is the Euclidean norm of vector $\overline{o_{\mathrm{i}}}$

2.2 Genetic Algorithm (GA)

Genetic algorithm has been successfully adopted in information retrieval. For instance, some studies have developed a ranking function discovery framework based on genetic programming, with various experiments performed to demonstrate how this novel framework can help to automate the ranking function design/discovery process [9]. Some studies have optimized parameters independently of retrieval models, thus enabling the optimized parameter set to illustrate the characteristics of the target test collections [10]. In the work, a genetic algorithm is used as optimization methods where the most difficult retrieval parameters to optimize are the feedback parameters owing to the lack of principles for calibrating them. The approach optimizes feedback parameters and basic scoring parameters simultaneously.

3 The Proposed Search of Topic

3.1 Proposed Search of Topic

The steps of proposed search of topic are explained as follows:

1. Data collection from corpus: We collect data from European Journal of Radiology, including the title, abstract, and author-defined keywords of each paper.

2. Document presentation with vector space model: The author-defined keywords are used as the feature candidates. First of all, the *tf-idf* value of each feature is calculated, and features with the value lower than the threshold (0.5*log(total number of papers)*0.95) are deleted. The threshold is obtained through many experiments and is verified to produce satisfactory results. Finally, each paper is represented by the base vector formed by the *tf-idf* values of the remaining features.
3. Initialization of the number of topics in each chromosome: The number of topics included in a chromosome (a topic selection) is initialized here. The initial number is set to $\frac{1}{4}\sqrt{\text{count of total papers}}$. The square root is used since the initial number will be still adequately small despite the fact that the number of published papers in a certain period increases quickly. Also, with a trial and error method, the setting is verified to be proper.
4. Initialization of the topic selections for GA process: The first generation of topic selections is initialized at random. A generation includes 20 chromosomes, and each chromosome stands for a topic selection including several topics.
5. Genotype converting of topic selection: Genes in a chromosome are decoded to determine which topics should be chosen. A topic is chosen as the value of its corresponding gene is "1".
6. Computation of relatedness among topics and features based on VSM: Relatedness among topics and features is computed using the cosine method.
7. Independent chi-square test of topics: Pairwise comparison of the independence between two topics is performed using the independent chi-square method.
8. Evaluation of fitness values of topic selections: The fitness value of each topic selection is calculated. Larger value indicates higher differentiation of the topic selection and more papers classified into the topics of this topic selection.
9. Termination criterion of genetic algorithm for topic selection: The termination criterion is evolution of 100 generations. If the criterion is met, terminate the search and then go to step (11).
10. Process of genetic algorithm for topic selection: In the selection process, from 20 chromosomes, 10 with the highest fitness values are selected and duplicated using the roulette wheel selection method. In the crossover process, double-point crossover method is adopted. In mutation, the mutation rate is defined as 2%, and the process is redirected to step (5).
11. Termination criterion for number of topics in a chromosome: The termination criterion for number of topics in a chromosome is set to $\sqrt{\text{count of total papers}}$. Experimental results indicate that the average count of papers classified into the selected topics is insufficient when the count of the selected topics is very high. Notably, a representative search of topicwith those topic selections cannot be found. This setting of termination criterion is used since the value is adequately large with trials and experiments. If the criterion is met, terminate the process and go to step (13).
12. Increment of number of topics in a chromosome: The number of topics in a chromosome is incremented by one. Go to step (4).

13. Determination of the final topic selection: The topic selection with best fitness value is considered first, but the average number of papers classified into topics of this topic selection must be in the top five. Otherwise, the topic selection with next best fitness value will be considered, and so on.
14. Search of topic: Apply the finally selected topic selection and the related features with relatedness degree beyond the threshold to present the main structure of journal papers in study.

3.2 Fitness Function Design for the Topic Selection Search

We design a fitness function used in genetic algorithm to search for representative topic selection consisting of several topics. To obtain more representative topics, two key factors are considered, discrimination between selected topics and the number of papers classified into selected topics. Therefore, the fitness function for a topic selection is shown as follows:

$$Fitness\ function = \sum_{i=1}^{I} \left(chi - square_i \right) * Avg_d \tag{3}$$

where I is the total number of topics, $chi\text{-}square_i$ is the value of chi-square test of topic I, and Avg_d is the average number of papers classified into selected topics.

4 Experiments

The data of European Journal of Radiology is used as target. The collected data items include titles, abstracts, and author-defined keywords. The sample is composed of papers published from issue 1 to issue 12, volume 81, 2012. During the period, 831 papers have abstracts, and author-defined keywords.

4.1 Presentation of Search of Topic

We attempt to construct search of topic to explore the major topics and related features of the papers published in a period. This search of topic is expected to help readers understand not only the major research topics as well as the related techniques and issues but also the research tendencies among the published papers.

Through information retrieval technique, genetic algorithm, and independent chi-square test, topics with higher representativeness are derived. With the selected topics and their related features, a search of topic is constructed. The topics and related features are from author-defined keywords. A search of topic that encompasses the topics of published papers, number of papers classified into each topic, and the related features (related techniques, methods, or issues) of each topic will be built.

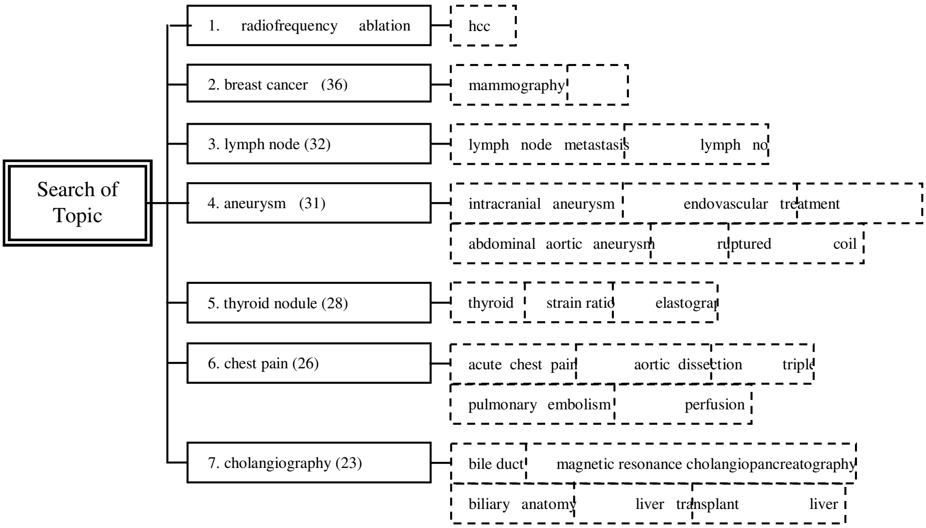


Fig. 1. Search of Topic for European journal of Radiology based on 2012 corpus

The search of topic is shown in Fig. 1. The first level of the proposed search of topic consists of the selected topics. The second level consists of features related to some of the selected topics. The average numbers of papers classified into the selected topics are 30.4 (the count within the parenthesis behind each topic is the number of papers classified into the topic).

4.2 Evaluation of Search of Topic

The search of topic is evaluated in aspects of topic representation, and appropriate paper classification:

1. *Evaluation of topic representativeness*：In the knowledge structure, the number of papers classified into each selected topic is clearly revealed. The ratio of the papers classified into the selected topics to the total number of papers is 26% (213/831).
2. *Evaluation of appropriate paper classification*：This is evaluated by the threshold of relatedness and the average relatedness between related features and topics. In terms of the threshold of relatedness, we set a threshold of 0.3 to determine whether a feature should be classified into a certain topic. All related features close to this threshold were searched for, and the abstracts of the papers containing these related features were analyzed artificially. We discovered that all these papers have been properly classified into topics.

5 Conclusions

The research results showed that search of topic constructed with the aid of genetic algorithm featured a certain degree of representativeness. Also, the average relatedness values in the two periods were higher than the threshold, indicating that all the papers should have been properly classified.

Search of topic was built in a completely automated manner, without any human interference. Unavoidably, the constructed search of topic could not process different keywords with similar meanings as the same keywords. However, the effect of our search of topic was not affected much.

For the future study, the follows can be considered.

1. In the study, two features with similar meaning are treated as different items. Expert knowledge may be employed to preprocess selected features. It might increase the precision of search of topic but it will certainly decrease the characteristics of automation and flexibility.
2. All journals in a specific area can be selected for analysis, so as to derive more meaningful implications. Therefore, future researchers are advised to analyze multiple journals in a specific area and compare with other approach.

References

1. Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., Gummadi, K.: Cognos: Crowdsourcing Search for Topic Experts in Microblogs. In: 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 575–590. ACM, New York (2012)
2. Amolochitis, E., Christou, I.T., Tan, Z.-H., Prasad, R.: A heuristic hierarchical scheme for academic search and retrieval. Information Processing and Management 49, 1326–1343 (2013)
3. Newman, D., Baldwin, T., Cavedon, L., Huang, E., Karimi, S., Martinez, D., Scholer, F., Zobel, J.: Visualizing search results and document collections using topic maps. Web Semantics: Science, Services and Agents on the World Wide Web 8, 169–175 (2010)
4. Baglivo, M., Winklhofer, S., Hatch, G.M., Ampanozi, G., Thali, M.J., Ruder, T.D.: The rise of forensic and post-mortem radiology-Analysis of literature between the year 2000 and 2011. Journal of Forensic and Imaging 1, 3–9 (2013)
5. Bradley, D.A., Wells, K.: Biomedical applications reviewed: Hot topic areas. Radiation Physics and Chemistry 95, 191–201 (2014)
6. Wetter, O.E.: Imaging in airport security: Past, present, future and the link to forensic and clinical radiology. Journal of Forensic Radiology and Imaging 1, 152–160 (2013)
7. Desjardins, B., Hamilton, R.C.: A Practical Approach for Inexpensive Searches of Radiology Report Databases 14, 749–756 (2007)
8. Sarkar, I.N.: A vector space model approach to identify genetically related diseases. J. Am. Med. Inform. Assoc. 19, 249–254 (2012)
9. Fan, W., Gordon, M.D., Pathak, P.: A generic ranking function discovery framework by genetic programming for information retrieval. Information Processing and Management 40(4), 587–602 (2004)
10. Fujita, S.: Retrieval parameter optimization using genetic algorithms. Information Processing and Management 45(6), 664–682 (2009)

A Study of String Matching System Based on Database Set Operation

Ming-Hsiung Ying and Chien-Yu Lin

Chung Hua University, Department of Information Management, HsinChu, Taiwan
mhying@chu.edu.tw, zw1018242000@gmail.com

Abstract. In recent years, people can easily get various data and information through Internet. People can copy the entire downloaded data, digitized information into own paper work, and form some plagiarism problems. Previous studies use of statistics, vectors matrices to compare string among documents. When someone change the location of words, and add some superfluous words or sentences between strings, it will be greatly reduced the accurate rate of matching system; Moreover, it may cause students keep plagiarism if matching system cannot find the alignments correctly. This study uses Chinese Word Segmentation and Database Set Operation as a base to construct a string matching system to solve the excessive superfluous words and order problems. Database Set Operation may be more efficient than the program with lots of words inside its memory. This study creates a prototype system, and the result of the prototype shows that the accuracy performance is performed well.

Keywords: String Matching, Database Set Operation, Chinese Word Segmentation, Matching System.

1 Introduction

In recent years, the Internet and information technology is very popular. People can easily get all kinds of data, information and knowledge through the Internet and information technology. Therefore, more and more people access to information and knowledge ways, gradually changed from traditional books, documents, etc., into the way through information technology. Many students copy the entire downloaded data, digitized information into own paper work, and form some plagiarism problems. In recent years, many schools appeared plagiarism issues. Many students' homework have copied and revised from other students' homework, and the teachers must to spend a lot of energy to match the homework content among students.

Previous studies use of statistics, vectors matrices to compare string among documents. When someone change the location of words, and add some superfluous words or sentences into the documents, it will be greatly reduced the accurate rate of matching system.

If matching system cannot find the alignments correctly, students will keep the plagiarism behavior. To solve the problem that adding some superfluous words and change the word order in documents, the string matching system cannot find the

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_43, © Springer-Verlag Berlin Heidelberg 2014

plagiarism words correctly. This study uses Chinese Word Segmentation and Database Set Operation as a base to construct a string matching system. This study will creates a prototype system, which will compare with methods of matching words, and T brand's matching system, to verify its efficiency and accuracy.

In particular, this study has the following objectives:

- The past string matching method cannot find some plagiarism words in the documents. Especially when the superfluous words have added into sentence or the words order have changed. This study uses Chinese Word Segmentation and Database Set Operation as a base to construct a string matching system to solve the excessive superfluous words and order problems.
- This study will compare with methods of matching words to verify the efficiency and accuracy of prototype system.
- The results of this study provide a reference for researchers in related fields, and hope to contribute in the field of string matching.

2 Literature Review

2.1 Extensible Markup Language

Extensible Markup Language (XML) was published in 1988 by the WWW Development Association (World Wide Web Consortium, W3C). XML and HTML is similar, the HTML has less expandability and flexibility than XML, and the XML has good scalability and class sex [1]. XML has not fixed format, and it is not a single language or defined description language, it is a Meta-Language [2]. In Figure 1, the XML file contains the data and structure, and it has become the standard for electronic data storage and exchange [3].

```
▼<toplevel>
  ▼<CompleteSuggestion>
     <suggestion data="中華大學資管系"/>
     <num_queries int="709000"/>
   </CompleteSuggestion>
 </toplevel>
```

Fig. 1. Figure 1 XML standard format

2.2 Chinese Word Segmentation

Lin et al. (2010) indicated the Chinese word segmentation is very important and basic when dealing with the analysis of Chinese documents , Chen(1992) also indicated that there are great differences in the structure and semantics of Chinese and English, the analysis of Chinese sentences single word may not be the smallest unit, for each word in terms of the English sentence is individually separated by a space[4].

In Chinese documents, to analyze the meaning of the string section, it must to execute the Chinese word segmentation processing, and then analyze the content of the vocabulary and comparison, to truly understand the words in the string to be expression meaning.

2.3 String Matching Method

Information is very important in today, there is a lot of information in the network, finding a particular substring in a large number of text is a big problem. Wu et al. (2008) indicated that the use of the keyword method in the past for comparison, but if the use of synonym or other word to replace the original word or string will result in inaccurate punctuation in the string. and in the past have made way to quickly search many scholars use different methods, such as: brute force search method, quick search method, etc. [5].

3 Research Methods and Design

3.1 The Superfluous Words Analysis

Many previous studies use statistics, vectors, matrices, and other methods to shift the position, and propose many different methods to compare string among documents. For example, in quickly search method, if adding a superfluous word "的" in BP, and the content is changed "激勵學生的參與及互動讓學生學習更加快速", the comparison with the process of adding superfluous words is shown in Figure 2.

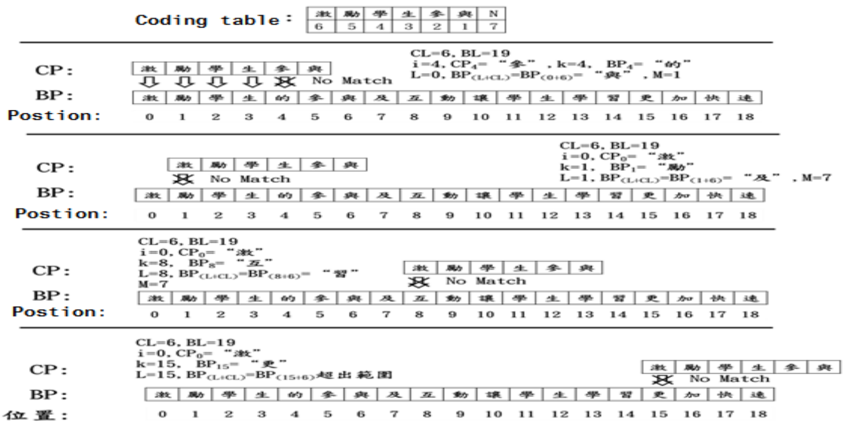


Fig. 2. Comparison with the process of adding superfluous words

Figure 2 shows the superfluous word into the string for comparison, will lead to the past compare method cannot find the specific string. Therefore, this study using the Chinese word segmentation and database set operations, proposed a new string match method to solve the problems of adding superfluous words and to change the word order.

3.2 Comparison Process

Figure 3 shows the string comparison process, the process is described as follows:

1. In this study, the prototype system will send some string to CKIP Chinese Word Segmentation System.

2. When the CKIP system has received the segmented results, our system will combined with Google XML, execute an online word repair process, in order to increase the accuracy of subsequent string matching.
3. The repaired words has stored into database, and mark the position of each word in the string, to provide subsequent string comparison.
4. Utilizing the set operations provided by the database to calculate the similarity between each string and other string, and superfluous words and punctuation are not included in the calculation.

If the word overlap ratio exceeds the threshold set by the system will to execute a string comparison process, the string comparison is to use the set operations of database. In this study, the default threshold is 20%, this threshold rate can revise by users.

3.3 Chinese Word Segmentation

The CKIP Chinese word segmentation system was developed by Academia Sinica of Taiwan. In this study, the prototype system will sent string to the CKIP Chinese word segmentation system to get the segmented words. However, the CKIP Chinese word segmentation system can receive a restricted number of words only. Therefore, this study has designed a sending string mechanism in batch, each time only send 250 words. The batch sending process is described as follows:

1. To determine whether a string more than 250 words, if the string more than 250 words, the string are divided and sent those words in batches, otherwise directly sent all words to the CKIP Chinese word segmentation.
2. If the string exceeds 250 characters, the system will automatically find the last comma from the front of 250 words, and then send the string before the comma to CKIP Chinese word segmentation system.
3. To determine whether there have remaining string has not been processed, if all strings have been processed, the prototype system to end the segmentation process. If there have some unprocessed string, then repeat the first step until all strings are processed.

3.4 Online Word Repair Process

Previous about the word repair studies, they use the built-in word-based to repair words. Now, every day may emerge some new word. Building an artificial word-based will spend a lot of manpower and time. This study using the online Google XML provided for online word repair process, without prior built-in word-based. The Figure 3 is a result of similar words found in Google XML, the Figure 4 is a result of similar words not found in Google XML. And the Figure 5 is a result of online word repair.

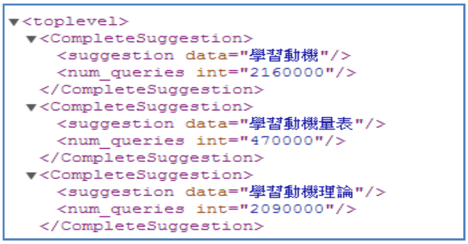


Fig. 3. The similar words is found in Google XML

Fig. 4. The similar words is not found in Google XML

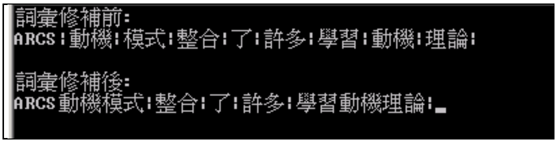


Fig. 5. The result of online word repair

3.5 Word Sequence Encoding and Calculating the Word Overlap Rate and Word Similar Rate

In Chinese string sequence between words is very important, if the order does not mean it is expressed may differ, so this research will encode a string sequential number for each word and stored those sequential number in the database. In this study, the overlap rate is set to 20%, if the overlap rate is greater than the 2-%, the prototype will start the string matching process. Formula 1 is the overlap rate formula

$$M = \frac{\mathrm{COUNT}(A \cap B)}{\mathrm{COUNT}(A)} * 100\% \tag{1}$$

M: word overlap rate
COUNT(A∩B): A number of the intersection with the B words
COUNT(A):The number of all vocabulary in A

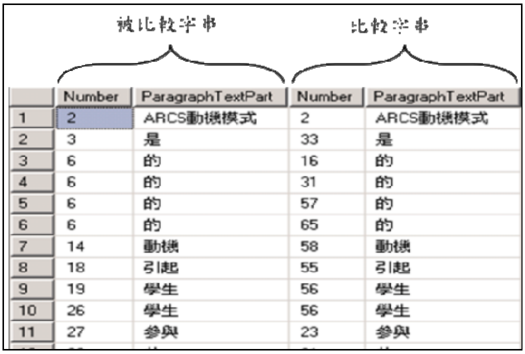
The similar rate refers to the same text number ratio that some text strings A is found in the string Y, the similarity rate is calculated as follows:

$$SR = \frac{\mathrm{COUNT}(\mathrm{CPsc})}{\mathrm{COUNT}(\mathrm{CPs})} * 100\% \tag{2}$$

SR: similar rate
COUNT(CPs): The total number of words comparing strings
COUNT(CPsc): The number of labeled words of comparing strings

3.6 Intersection Operator

Use the intersection operator provided by database,the comparing strings intersect the being compared string, in the intersection operation is not the punctuation into them, the sort order is by comparison string (Figure 6).



	Number	ParagraphTextPart	Number	ParagraphTextPart
1	2	ARCS動機模式	2	ARCS動機模式
2	3	是	33	是
3	6	的	16	的
4	6	的	31	的
5	6	的	57	的
6	6	的	65	的
7	14	動機	58	動機
8	18	引起	55	引起
9	19	學生	56	學生
10	26	學生	56	學生
11	27	參與	23	參與

Fig. 6. Intersection and Ordering

4 Verify the Effectiveness of the System

In the end, the result of the prototype shows that the accuracy performance is 100 % identified and highlighted the same wordings in strings. The result of accuracy ratio is shown Figure 7.

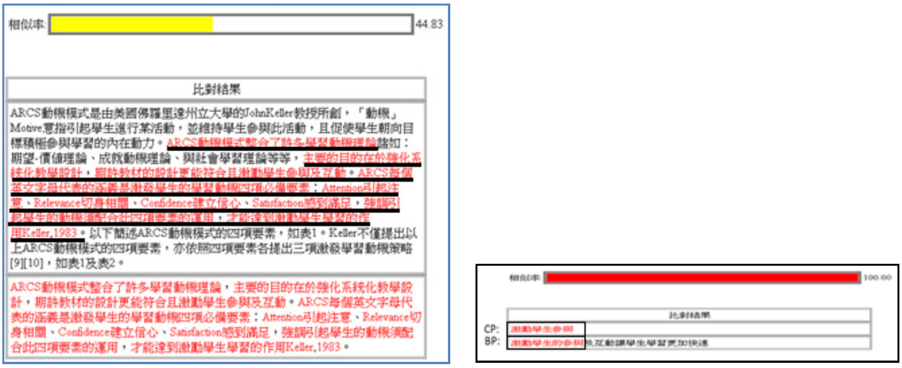


Fig. 7. The result of accuracy ratio

References

1. Lee, J.T.: Development of XML-Based Geo-Spatial Information Distri-bution System. Journal of Cartography 16, 191–204 (2006)
2. Maarouf, M.Y., Chung, S.M.: XML Integrated Environment for Service-Oriented Data Management. In: IEEE International Conference on Tools with Artificial Intelligence, vol. 2, pp. 361–368 (2008)
3. Zhang, J., Lang, B., Duan, Y.: An XML Data Placement Strategy for Dis-tributed XML Storage and Parallel Query. In: 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 433–439 (2011)
4. Lin, C.X., Chen, Z.J., Ling, C.C.: Combined with a long term priority sequence labeled with Chinese word segmentation research. Information Security Communications 15(3-4), 161–179 (2010)
5. Wu, D., Zhou, X., Zhang, H.: The Pattern Matching Algorithms Formalized Analyze in Chinese Strings. Intelligent Information Technology Application 1, 403–407 (2008)

A Study of Intention of Classroom Interactions in IRS-Equipped Learning Environment

Kuan-Hsi Lin[1] and Shiang-Lin Lin[2]

[1] HABOOK Information Technology Inc. Training Specialist, R.O.C.
francis@aclass.com.tw
[2] Department of Management Information Systems,
National Chengchi University, Taipei, R.O.C.
shiang0623@gmail.com

Abstract. In nowadays society, the most common manner that students recevice education is through teachers giving lectures in the classroom. In Taiwanese schools, the lecture method that teachers play a role of a knowledge dispenser instead of a learning facilitor is commonly adopted. Since late 90's, however, the educational technology has become gradually popular in teaching enviroments. The aim of this study is to the Taiwanse college students' intention of classroom interactions through the IRS (Interactive Response System) devices. The results show that learners' goal commitment and perceived usefulness are affected by an IRS-equipped learning environment which leads to the influence of learners' intention to ineteract and participate in classroom activities.

Keywords: Educational technology, Interactive Response System, Innovation Diffusion Theory, Social Cognitive Theory.

1 Introduction

In nowadays society, the majority of students in most countries receive education in the school through lecturers. Knowledge is delivered to learners by professional lecturers through the textbooks and various teaching resources.

Traditionally, the educational theories of instructionism that emphasize teacher behavior are commonly described and applied in knowledge delivering (Jonassen, 1991; 1995). On the other hand, based on the theories of social constructivism on knowledge receiving, the learning process can be influenced by the society and environment. In addition, verbal expressions are the major medium in learning process for community interaction and knowledge delivering (Vygotsky, 1986). The theories of social constructivism not only emphasize the importance of learning environment but the significant role of verbal interaction in the enhancement of learning quality. Hence, to improve the teaching quality, learners should be strongly encouraged to communicate more with each other as well as the lecturer and receive knowledge from the interaction (Kern, 1995).

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_44, © Springer-Verlag Berlin Heidelberg 2014

With the advent of the Information Age and improvements in IT, the integration of education and technology that aims to improve learning progress has become a significant issue that is broadly discussed.

For the past decade, many IT companies have invested in the field of educational technology that promote the teaching strategies using technology. The IRS (interactive response system) innovative teaching developed by HABOOK Inc, for example, aims to the application of technical devices such as interactive whiteboard and IRS to improve the classroom interactions and learning outcomes. The IRS is current commonly adopted in Taiwanese schools (Wu, et al., 2013). This system allows students to answer questions on a remote control and displays the answer analysis instantly in the classroom (Wu et al., 2011). In addition, the feature of giving answers anonomously reduces the fear students have for being embarrassed by answering questions incorrectly.

Due to the popularization of e-learning, studies and researches related to curriculums using educational technology including applying IRS in learning environment are widely conducted (Liu et al., 2003；Liu et al., 2008；Zhuang et al., 2011). Yet, the learner’s intention to classroom interaction was rarely discussed. Issues on using IT devices and software to motivate learners and improve teaching are required to be discussed deeply. Since the application of IRS in Taiwanese higher education is fast developing, this study focuses on the learner’s intention on participation and interaction in colleges affected by IRS teaching strategies.

2 Research Hypothesis

In his research, Bandura (1986) indicates that human behavior is depicted as being shaped and controlled by environmental influences or internal dispositions. He also points out that the cognitive process is a major factor that influences human behaviors as well as the beliefs and attitudes that serve to motive self-efficacy. An individual’s self-efficacy is found that be a significant factor in how goals and tasks are approached. In the goal-setting process, the participation enhances goal commitment which motivates people to accomplish tasks (Locke & Latham, 1990).

Taylor & Todd (1995) suggest in their research that consumers’ purchase intentions can be improved by the confidence and expectation on the relative advantages of an innovative product. When consumers figure out an innovative product or service coming with relative advantages along with less compatibility and complexity, their self-confidence is improved and therefor the adoption of the product grows (Liao et al., 1999). According to the arguments above, this study proposed three hypotheses:

H1. Innovation of learning environment have a significant positive influence on self-efficacy of learning.

H2. Self-efficacy of learning have a significant positive influence on goal commitment of learning.

H3. Goal commitment of learning have a significant positive influence on intention of interaction.

In his research, Rogers (1983) points out the cognitive theory play a significant role in innovation diffusion; such phenomenon can also been seen in various fields. With higher innovative feature comes with higher consumer acceptance and increased positive behavior intention (Citrin et al., 2000). The innovation diffusion is major factor that affects an individual's expectation and acceptance in innovative decision making (Gera & Chen, 2003). In addition, users' behavioral intentions are impacted by the perceived usefulness (Davis, 1989). Based on those theories, two more hypotheses are proposed in this study, and the hypothesis model is displayed in Fig. 1.

H4. Innovation of learning environment have a significant positive influence on perceived usefulness.

H5. Perceived usefulness have a significant positive influence on intention of interaction.

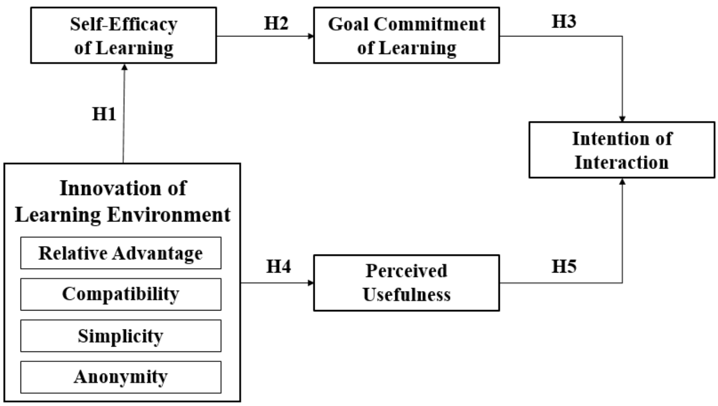


Fig. 1. The Hypothesis Model

3 Theoretical Background and Research Methodology

According to Rogers' research on diffusion of innovations, innovations are the key points for users to adopt new technology (Rogers, 1995). Based on the statement, this study brings the concept of innovations into the model and develops the construct of "innovations in teaching methods that focuses on the effects of the IRS applications in an innovative teaching environment on learners' intention of interaction.

Rogers also indicates that the variance of innovation adoption is explained by relative advantage, compatibility, complexity, trailability, and observability. Since this study focuses on the innovations in teaching methods instead of a commercial product, the variance of trailability will not be discussed in this study. The relative advantage in this study is defined as the relative advantage of innovative teaching methods using IRS in class compared to traditional teaching methods. The compatibility indicates the actual need for learners to use IRS in favor of their learning. The complexity is referred to learner's understanding on using IRS in class interactions. The answers given by students through IRS can be only viewed by the lecturers. In other word, it provides the anonymity for students when they are encouraged to respond and participate in class activities which motivate them to express their own opinions without being influenced by peer pressure.

Self-efficiency, according to Bandura (1986) provides faith and confidence to an individual to fulfill and accomplish particular goals and attract a goal directed activity. By making self-satisfaction conditional on matching adopted goals, people give direction to their behavior and create incentives to persist in their efforts until they fulfill their goals. Based on his statement, this study defines self-efficiency as the learners' confidence on participating in a class discussion and interacting with the teachers using IRS device in the innovative teaching environment. The goal commitment is defined as learners' expectation on classroom interaction and innovative teaching strategies.

Venkatesh & Davis (2000) explained that perceived usefulness is the degree to which a user believes that using technology would enhance his or her performance while others (Kwon & Wen, 2010) suggested it is users' awareness on his or her performance improved by information technology. In this study, perceived usefulness is defined as learners' awareness of their motivation for class interactions. The intention of interaction measured in this study refers to the degree of learners' willingness to interact in a class using an IRS device.

The scale of measurement of this study is developed based on the theories and criteria described above and shown in Appendix A.

4 Research Analysis

The participants of the study were college and graduate school students in Taiwan who were introduced the IRS and innovative teaching in a real class through a video demonstration。208 valid questionnaires were returned out of 212. 48% of the questionnaires were done by males and 52% females. More than half (56%) of the participants were under 20 years old. Most of the participants (77%) are enrolled in public schools. The table of demographics is displayed in Table. 1.

Table 1. Demographic Profile of Respondents

Variable	Sub-variable	Number	Percent
Gender	Male	100	48%
	Female	108	52%
Age	Under 20	116	56%
	21~25	32	15%
	26~30	56	27%
	31~40	4	2%
Type of school	Public school	160	77%
	Private school	48	23%

4.1 Measurement Model

The Partial Least Squares (PLS) method was used to analyze the reliability and validity of the research model. According to Kerlinger (1999), the reliability analysis measures something consistently and repeatedly. The Cronbach`s α is commonly used as an estimate of the reliability and a satisfactory level of reliability is an alpha of 0.6 or above (Hair et al., 1998). The Cronbach`s α of all constructs in this study is higher than 0.6, which indicates the research model is relatively reliable.

The validity of the research model is measured by convergent and discriminate validity which composite reliability and average variance extracted (AVE) values were examined in Table 2. Fornell & Larcker (1981) suggested that composite reliability should be great than 0.6. The average variance extracted should be greater than 0.5 for the sake of convergent validity.

Table 2. Reliability and Validity Analysis

Construct & dimension	Cronbach`s **α**	CR	AVE
Relative advantage (ra)	0.81	0.89	0.72
Compatibility (co)	0.78	0.87	0.69
Simplicity (si)	0.83	0.90	0.75
Anonymity (an)	0.83	0.90	0.7 4
Self-efficacy of learning (SE)	0.87	0.91	0.72
Goal commitment of learning (GC)	0.90	0.93	0.72
Perceived usefulness (PU)	0.90	0.93	0.83
Intention of interaction (INT)	0.60	0.93	0.82

Fornell & Larcker (1981) also suggest the square root of the AVE should be greater than the correlation shared between the construct and other constructs in order to demonstrate satisfactory discriminant validity. All of the correlations in this research model meet this criterion. The correlations matrix is shown in Table 3.

Table 3. Correlations of Latent Variables

	GC	an	co	si	ra	INT	PU	SE
GC	**0.85**							
an	0.48	**0.86**						
co	0.64	0.53	**0.83**					
si	0.32	0.17	0.33	**0.87**				
ra	0.59	0.51	0.60	0.48	**0.85**			
INT	0.75	0.44	0.64	0.40	0.65	**0.91**		
PU	0.83	0.43	0.65	0.46	0.70	0.74	**0.91**	
SE	0.75	0.47	0.59	0.30	0.49	0.58	0.62	**0.85**

*Diagonal values are the square root of AVE.

4.2 Structural Model

Reaching satisfactory reliability and validity, the study uses PLS method to analyze each hypothesis in the research model. The path coefficient and the explanatory power of the model presented in Fig.3.

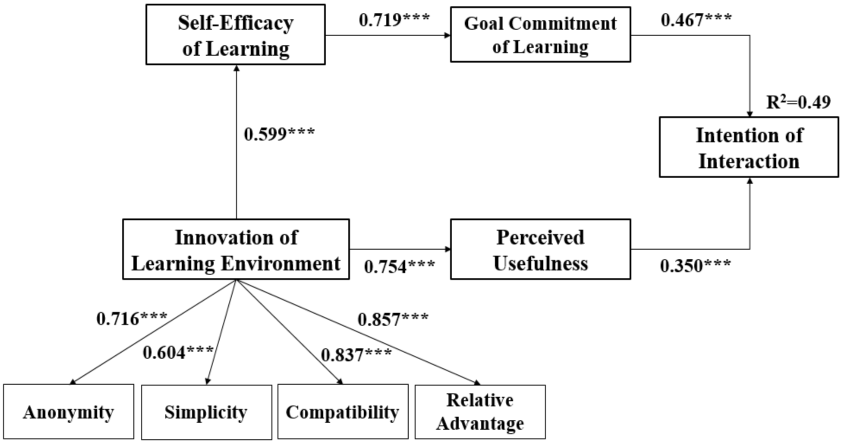


Fig. 2. Result of structural model

The results show that "self-efficacy of learning" is greatly affected by the "innovation of learning environment", which also influenced the intention of interaction" through learners' "goal commitment". The learners agree that the IRS device and the corresponding teaching styles are user-friendly and easy to follow. Through the device and strategies, multiple learners are able to respond to multiple-choice questions simultaneously and receive the results in real time, which helps learners realize how much they understand a lesson right after they are assessed. Moreover, the device and strategies allow learners to give answers to questions anonymously. That causes the reduction of fear for making a wrong answer and thus motivates learners to respond to a question.

In addition, the results illustrate that the "innovation of learning environment" influence "perceived usefulness" and eventually the "intention of interaction". Since the IRS-equipped innovative teaching environment provides a great opportunity for learners to interact with the teachers in class activities, students can express their ideas better and improve their learning outcomes.

In terms of the explanatory power, "the goal commitment of learning" is set for learners using IRS in an innovative teaching environment to improve learning outcomes while the "perceived usefulness" is the learners' awareness on the improvement of learning quality and general educational environment. These two constructs deeply influence learners' "intention of interaction" in an IRS-equipped environment. The research model accounts for 49% of variance explained.

5 Conclusions and Further Work

With the rapid development and popularization of information technology in the 21st century, transforming traditional teaching and learning though technology has become a widespread issue. Therefore, both teachers and learners' needs and intentions on using technology in a learning environment should be taken into account to maximize the educational benefit of technology in classrooms.

5.1 Managerial Implication

Academically, the study focuses on the improvement of college students‘ intention to interacte and participate in classroom activities based on an innovative teaching strategies using the device of IRS (interactive response system). The results specify college students gerenally agree they are motivated by the system and teaching strategies to interact and participate in class activities and discussions. Meanwhile, the mojority of colleage students are willing to enagage to class interactions and communication according to ther result of this study.

In practice, the results indicate that in an IRS-equipped learning environment, learners' motivation on giving opinions and participating in classroom activities is improved. As a result, it is suggested that teachers and educators import educational technology in their teaching environments and encourage learners to interact with the teachers using technical devices.

5.2 Research Limitation and Future Work

The participants of this study were Taiwanese college students that have never been introduced IRS and it's corresponding teaching strategies. Since the IRS is not widely quipped in Taiwanese colleges and universities, the participants in this study were led to experience the IRS applications and innovative teaching strategies through simulation scenarios. Further research on learners' experience of using IRS and related educational technologies to participate in an authentic classroom should be conducted to determine the impacts of technology in education.

References

1. Bandura, A.: Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall (1986)
2. Citrin, A.V., Sprott, D.E., Silverman, S.N., Stem, D.E.: Adoption of Internet Shopping: The Role of Consumer Innovativeness. Industrial Management and Data Systems 100(7), 294–300 (2000)
3. Davis, F.D.: Perceived Usefulness Perceived Ease of Use, and User Information Technology. MIS Quarterly 13(3), 319–340 (1989)
4. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable and Measurement Error. Journal of Management Research 18, 39–50 (1981)

5. Gera, K., Chen, L.D.: Wireless Technology Diffusion: An Evaluation Model for Wireless Services. In: Association for Information Systems (AMCIS Proceedings) (2003)
6. Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C.: Multivariate Data Analysis, 5th edn. Prentice-Hall (1998)
7. Jonassen, D.H.: Objectivism vs. Constructivism. Educational Technology Research and Development 39(3), 5–14 (1991)
8. Jonassen, D.H., Davidson, M., Collins, M., Campbell, J., Haag, B.B.: Constructivism and Computer-Mediated Communication in Distance Education. American Journal of Distance Education 9(2), 7–26 (1995)
9. Kerlinger, F.N., Lee, H.B.: Foundations of Behavioral Research, 4th edn. Cengage Learning (1999)
10. Kern, R.G.: Restructuring Classroom Interaction with Networked Computers: Effects on Quantity and Characteristics of Language Production 79(4), 457–476 (1995)
11. Kwon, O., Wen, Y.: An Empirical Study of the Factors Affecting Social Network Use. Computers in Human Behavior 26, 254–263 (2010)
12. Liao, S., Shao, Y.P., Wang, H., Chen, A.: The Adoption of Virtual Banking: An Empirical Study. International Journal of Information Management 19(1), 63–74 (1999)
13. Liu, T.C., Liang, J.K., Wang, H.Y., Chan, T.W., Wei, L.H.: Embedding Educlick in Classroom to Enhance Interaction. In: Proceedings of International Conference on Computers in Education (ICCE), pp. 117–125 (2003)
14. Liu, T.C., Lin, Y.C., Bhattacharya, M.: Introducing Learning Technologies into Classroom in Accordance with Teacher's Instructional Approach. In: Advanced Learning Technologies (2008)
15. Locke, E.A., Latham, G.P.: A Theory of Goal Setting and Task Performance. Prentice Hall (1990)
16. Rogers, E.M.: Diffusion of Innovation, 4th edn. Free Press (1995)
17. Rogers, E.M.: Diffusion of Innovations, 3rd edn. Free Press (1983)
18. Taylor, S., Todd, P.A.: Decomposition and Cross Effects in the Theory of Planned Behavior: A Study of Consumer Adoption Intentions. International Journal of Research in Marketing 12, 137–155 (1995)
19. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 46(2), 186–204 (2000)
20. Vygotsky, L.S.: Thought and Language. MIT Press (1986)
21. Wu, C.W., Chang, Y.H., Shu, C.M., Wu, T.C., Wang, S.Y.: Smarter Classroom and Innovative Technology: Theory and Cases. AClass (2013)
22. Wu, C.W., Han, C.T., Wang, S.Y., Wu, T.C., Chang, Y.H.: Instructional Technology and Innovative Teaching: Theory and Cases. AClass (2011)
23. Zhuang, Z.Q., He, J.M.: The Design and Application of IRS-Based Classroom Assessment System. Modern Educational Technology 1(32) (2011)

Software Recommendation of Application Market by Semantic Network and Multi-Criteria Decision Analysis

Chih-Kun Ke[1], Shih-Fang Chang[2], Chang-Yu Jen[1], and Jia Chi Liao[1]

[1] Department of Information Management,
National Taichung University of Science and Technology,
40401 Taichung, Taiwan R.O.C.
[2] Industrial Technology Research Institute, Intelligent Computing Platform Div.,
Computational Intelligence Technology Center, Hsinchu, Taiwan R.O.C.
{ckk,s1801B104,s1802B110}@nutc.edu.tw,
finney@itri.org.tw

Abstract. In software as a service (SaaS) model, application market provides various software services for users to access. However, large amount of software services are difficult to choice because of few attributes and incomplete description to illustrate their functionality. Besides, the fetch results from application market may not match user preference and waste user much time to get the desired software service. In this work, we propose an approach to improve software service searching effectiveness in an application market. Several advanced techniques are enforced. Information retrieval technology analyzes the description of a software service to get its key concepts. The association rule mining technology discovers the hidden association between various software service key concepts. The relationships of software service key concepts and discovered association rules are built a semantic network to connect relevant key concepts of software services. After configuring the software service attributes for quality of service consideration, the multi-criteria decision analysis is used to get the ranking order of the candidate software services. The software services key concepts, discovered association rules, semantic network, and multi-criteria decision analysis approach are built a recommendation system. User gets the reasonable software service based the ranking order of candidates from the recommendation system. We hope the proposed approach facilitates user to get the software service effectively in a popular application market.

Keywords: software as a service, application market, semantic network, multi-criteria decision analysis, software recommendation.

1 Introduction

Application market [1], e.g., Apple store, Google Play, Windows Marketplace of Microsoft, provides various software services (Apps) for user to access. Huge amount of software services are produced and promoted day by day, e.g., over one million software services are appeared in Google Play nowadays. User access an application market to get software services promoting mobile commerce for society, education,

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_45, © Springer-Verlag Berlin Heidelberg 2014

entertainment, and job, etc. However, an application market provides incomplete description and few attributes of a software service to illustrate its functionality. User may waste much time to search a software service and not easy to select a desired one. Getting a desired App from huge amount of software services in an application market becomes a challenge. The fetched software service which matches a user preference is also a key concern to keep the user loyalty of a specific application market.

Incomplete description of a software service requires more relevant information to reinforce its functionality representation. Information retrieval technology [2] analyzes the semi-structure of the incomplete description to get the key concepts of a specific software service. A semantic network is built to connect the key concepts of various software services. Besides, some relationship may hidden between the key concepts of various software services. Association rule mining [3, 4] discovers the hidden relationship between the key concepts of various software services. The discovered knowledge rules are used to reinforce the semantic network to construct more key concepts connection [5]. Attributes of software services are the clues for a user to search and get a desired software service. Based on the quality of service (QoS) consideration, a user may evaluate some attribute value as criteria to select a reasonable software service from various candidates. According to the various issues of interest, multi-criteria decision analysis (MCDA) helps user to select a reasonable software service from huge number of candidate software services [6]. A multi-criteria decision analysis is an approach which structures and solves the multi-dimensional and complicated problems, especially the decision and planning problems involving multiple criteria. Generally speaking, such problems do not own a unique optimal solution, so a decision maker needs to use his/her preferences to differentiate between various solutions. Therefore, a multi-criteria decision analysis approach formulates the selection order of the various candidate software services to optimize the user's ability to get a reasonable software service [7].

This work presents an approach to improve software service searching effectiveness in an application market. Several advanced techniques are enforced. Information retrieval technology analyzes the incomplete description of a software service to get its key concepts. The association rule mining technology discovers the hidden relationship between key concepts of various software services. The key concepts of software services and discovered association rules are built a semantic network to connect relevant key concepts of software services. After configuring the software service attributes for quality of service (QoS) consideration, the multi-criteria decision analysis is used to evaluate the ranking order of the candidate software services. The key concepts of software services, discovered association rules, semantic network, and multi-criteria decision analysis approach are built a recommendation system. User searches and gets the reasonable software service based the ranking order of candidates from the recommendation system.

The rest of this paper is organized as follows. Section 2 reviews pertinent literature on cloud service, application market, association rule mining, and multi-criteria decision analysis. Section 3 then introduces the proposed approach to improve software service searching effectiveness in an application market. Conclusions are finally, drawn in Section 4, along with experiments for future research.

2 Related Works

This section discusses related technologies, including cloud service, application market, semantic network, association rule mining, and multi-criteria decision analysis.

2.1 Cloud Service and Application Market

Could computing [8] is a concept which reduces enterprise costs in equipment and preforms computing in a distributed network environment. Enterprise provides various cloud computing services in heterogeneous platforms for consumers to use the computing resources. The service models of cloud computing include Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). In Software as a Service (SaaS) model, a provider deploys and licenses software to customers for using as a service on demand. The software service may host on the SaaS software vendors' own servers. Customer downloads the software service to device, disable the software service after using or the on-demand contract expired [1].

Application marketplace [1] is a service which allows customers to explore and get software services from the vendors' servers. The famous and typical model for application marketplace includes Apple store, Google Play, and Windows Marketplace of Microsoft, etc. Developer uses software development kit (SDK) to design various software services. The accessible software services were published through marketplace's owner. Customer uses the specified client interface to get the software services directly to target device. The software service may charge by vendor or free to use.

2.2 Semantic Network and Association Rule Mining

Semantic network [5] is a kind of knowledge map to be a human knowledge representation model. Semantic network is used in artificial intelligent domain allowing computer to store, memorize, and learn human language knowledge. The knowledge ontology is used to define the key relationship between terms. The semantic network is constructed by nodes and edges. A node presents a concept or a term. An edge shows the relationship between nodes. Association rule mining [3, 4] is a method of data mining technology. It tries to find an association between item X and item Y from a large transaction database. The quality of an association rule is indicated by two measures, support and confidence. The support of a rule is the percentage of transactions that contain both item X and item Y. The confidence of a rule is the fraction of transactions containing item X that also contain item Y.

2.3 Multi-criteria Decision Analysis Method, ELECTRE

Multi-criteria decision making (MCDM) [6] explicitly considers multiple criteria in decision-making environments. It is a sub-discipline of operations research. Multi-criteria decision analysis is an approach which structures and solves the multi-dimensional and complicated problems, especially the decision and planning problems involving multiple criteria. Generally speaking, such problems do not own a unique optimal solution, so a decision maker needs to use his/her preferences to differentiate between various solutions. ELECTRE (Elimination Et Choice Translating Reality) [6, 7] is a family of multi-criteria decision analysis methods. ELECTRE methods involve two main steps. First step constructs the outranking relations for a comprehensive comparison of each pair of actions. Second step elaborates the recommendations from results obtained by an exploitation procedure from the first step.

3 The Approach for Improving Software Service Searching in an Application Market

This section introduces the proposed approach for searching software service in an application market, is shown in figure 1. The relevant information, e.g., attributes and description, of software services in an application market is collected and analyzed in data preprocessing. The key concepts of software services and association rules are constructed a semantic network to store in a knowledge base. Through the search interface, user inputs nature language description and configures the attributes based on QoS consideration to present his/her software service preferences. The input description is analyzed in data preprocessing to get key concepts. The key concepts are used as clues to semantic network to fetch the candidate software services [9]. The multi-criteria decision analysis approach evaluates the ranking order of the candidates to recommend user the top-k software services.

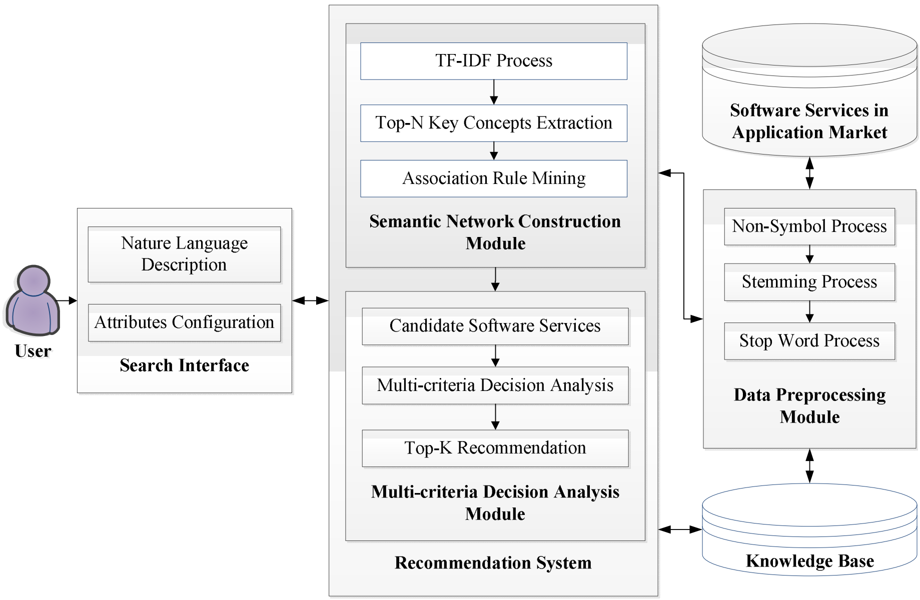


Fig. 1. An approach for improving software service search in an Application Market

The functionality of the proposed approach includes data preprocessing, semantic network construction, multi-criteria decision analysis, and knowledge base.

- **Data Preprocessing Module.** Web crawler technology [10] collects the relevant information of software service in an application market, including software service's attributes and description. Software service description is semi-structure text type which needs text pre-processing. During text pre-processing, the term transformation steps includes case folding, stemming, and stop word removal. Then, term weighting is applied to extract the most discriminating terms. [11-13]
- **Semantic Network Construction Module.** Semantic network (SN) construction module extracts the keywords and discovers the association rules between keywords. Keyword extraction process uses the term frequency and inverse document frequency (TF-IDF) concepts to calculate the word importance from preprocessed data. Then the top-n technology indicates top n words as the keywords (key concepts) to the specific document. The SN module analyzes the relationship between key concepts by association rules mining. The highest confidence key concepts are stored in a knowledge base and eliminated their relevant association rules. Other key concepts formed the association rules are used to construct the semantic network connection.
- **Multi-criteria Decision Analysis Module.** Multi-criteria decision analysis (MCDA) module compares the extracted keywords and semantic network to discover the software services which be similar with user requirement. User configures multi-criteria parameters of candidate software services and then MCDA module executes two steps to get the ranking order of candidate software services. First step constructs the outranking relations for a comprehensive comparison of each pair of actions. Second step elaborates the recommendations from the results obtained by an exploitation procedure from the first step. The ranking order is provided to the user for selecting the reasonable software service.
- **Knowledge Based Module.** The knowledge base stores the stop words and discovered knowledge rules from association rules mining. The set of stop words includes 421 words [14] and eliminated words. The eliminated words are the highest confidence keywords in a data set. The keywords represent the core concept of a software service. If the keywords with the highest confidence cannot use to represent the core concept of a software service. Therefore, the eliminated words should store in a set of stop words in knowledge base.

4 Conclusion

This work presents an approach to improve software service searching effectiveness in an application market. Several advanced techniques are enforced. Information retrieval technology analyzes the description of a software service to get its key concept. The association rule mining technology discovers the hidden association between various software service key concepts. The relationships of software service key concepts and discovered association rules are built a semantic network to connect relevant key concepts of software services. After configuring the software service attributes for quality of service (QoS) consideration, the multi-criteria decision analysis technology is used to get the ranking order of the candidate software services. The software services key concepts, discovered association rules, semantic network, and multi-criteria decision analysis approach are built a recommendation system. Then user gets the reasonable software service based the ranking order of candidates. In future work, we carry out experiments to evaluate the proposed approach to improve the effectiveness of software service search problem of application market. We hope the experimental results can show the proposed approach facilitating user to get the software service in a popular application market.

Acknowledgement. This research was supported in part by the National Science Council of Taiwan (Republic of China) with a NSC grant NSC 102-2410-H-025-017.

References

1. Chang, S.F.: A Reference Architecture for Application Marketplace Service Based on SaaS. International Journal of Grid and Utility Computing 2(4), 243–252 (2011)
2. Zhang, W., Yoshida, T., Tang, X.J.: A comparative study of TF*IDF, LSI and multi-words for text classification. Expert Systems with Applications 38, 2758–2765 (2011)
3. Agrawal, R., Imilienski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
4. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large database. In: Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
5. Jordy, S., Flavius, F., Frederik, H., Vadim, C.: Semantic Web service discovery using natural language processing techniques. Expert Systems with Applications 40, 4660–4671 (2013)
6. Ke, C.K., Chen, Y.L.: A message negotiation approach to e-services byutility function and multi-criteria decision analysis. Computers and Mathematics with Applications 64(5), 1056–1064 (2012)
7. Chi, Y.L., Lee, C.W., Chen, C.Y.: A Selection Approach for Optimized Web Services Compositions. Electronic Commerce Studies 2(3), 297–314 (2004)
8. Andrzej, G., Michael, B.: Toward dynamic and attribute based publication, discovery and selection for cloud computing. Future Generation Computer Systems 26, 947–970 (2010)
9. Ke, C.B., Huang, Z.Q.: Self-adaptive semantic web service matching method. Knowledge-Based System 35, 41–48 (2012)
10. Sotiris, B., Euripides, G.M.P., Evangelos, M.: Improving the performance of focused web crawlers. Data & Knowledge Engineering 68, 1001–1013 (2009)
11. Lovins, J.B.: Development of a Stemming Algorithm. Mechanical Translation and Computational Linguistics 11, 22–31 (1968)
12. Rob, H., Chris, P.: The Lancaster Stemming Algorithm, `http://www.comp.lancs.ac.uk/computing/research/stemming/Links/algorithms.htm`
13. Porter, M.: An algorithm for suffix stripping. Mechanical Translation and Computational Linguistics 11(1), 22–31 (1997)
14. Fox, C.: A Stop List for General Text. ACM SIGIR Forum 24(1-2), 21 (1989)

A Single Sign-On Mechanism with RBAC in Cloud Environment

Mei-Yu Wu[1], Shih-Fang Chang[2], and Yi-Jheng Wu[1]

[1] Department of Information Management, Chung Hua University, Hsinchu, Taiwan, R.O.C.
{mywu,M10010008}@chu.edu.tw
[2] Industrial Technology Research Institute, Intelligent Computing Platform Div., Computational Intelligence Technology Center, Hsinchu, Taiwan R.O.C.
finney@itri.org.tw

Abstract. With the rapid development of Internet, hardware performance and software technology had continuous improvement. Users have increasingly high demands for the Internet services. The concept of cloud services is derived slowly. Users could use a variety of different types of application services and access information regardless when and where through the Internet cloud. When using different cloud service, users may need a new registration or re-execute the login action. That is quite time-consuming and a repetitive action. Users maybe use different username and password to register in different cloud service. In this way, this increases the burden on user memory and managing their usernames and passwords. In order to avoid repeat register and login, this study proposes a mechanism for cloud environment that combining role-based access control (RBAC) authorization mechanism and single sign-on. In this architecture, users need not to have second login or a new registration. After the first login, user can use cloud services in the new domain. Complex works such as confirm and exchange information with the user data center are turned over to the cloud service provider.

Keywords: Cloud Services, Role-Based Access Control, Single Sign-On.

1 Introduction

Due to the advancement of technology, many users use laptops or smart phones and other mobile devices to connect the networks. Regardless of when and where the user is located, users can use a variety of cloud services through the Internet. Users do not have to worry about whether having adequate infrastructure and system resources to support the use of services. In contract, service providers will concern about user requirements and provide the enough resources and develop a number of different cloud services. This is a faster speed and more convenient service for users.

The users use various services that may be in different form of cloud such as private cloud, public cloud, and hybrid cloud. Different form of cloud do not exchange user information, even different service provider in the same form of cloud will not share user information with each other. As a result, before users want to use different

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_46, © Springer-Verlag Berlin Heidelberg 2014

cloud service, users must registration in each cloud service of each form of cloud environment. Each registration needs lengthy process and inputs complicated and similar basic data. Some cloud services need to wait the confirmed letter issued from service provider. The confirmed letter may delay several days. After successful registration, users must manage multiple accounts and passwords. While user uses different services at the same time, the user needs to enter account and password repeatedly. It is very time-consuming for users. The existing literatures provided several mechanisms to achieve the purpose of single sign-on [2][7][16][19]. However, users still need to complete the registration process of cloud services on different cloud providers in advance. Then users can use single sign-on service to use registered cloud services, but users still cannot use unregistered services. The research proposed a new architecture to solve the above problems under the premise of not changing the user experiences. In the proposed architecture, users do not need to study hard professional settings or learn new skills. Users use browsers to use cloud services according to the original habits. Users only need to register in a cloud service server before and successfully performed a login. Then the users can use registered cloud services or other unregistered cloud services whether in the public clouds, private clouds, or hybrid clouds. Besides, this study combined role-based access control mechanism to achieve efficient and maintained management.

2 Related Work

2.1 Cloud Service

According to the definition of National Institute of Standards and Technology (NIST), cloud computing is a model for enabling convenient, on-demand network access to a share pool of configurable computing resource that can be rapidly provisioned and released with minimal management effort or service provider interaction [9]. Cloud computing can be divided into cloud computing technologies and cloud computing services. Cloud computing technology focused on providing cloud computing capabilities, storage capacity and redundancy capabilities. Cloud services focused on providing services to or from the cloud. Users always cloud use any of the services, as long as connecting to cloud services providers with no additional equipment or cost [8][15].

Cloud computing providers offer their services according to several fundamental models, including infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The deployment models of cloud computing includes private cloud, public cloud, community cloud, and hybrid cloud [9][10][11][17].

2.2 Access Control

The main purpose of access control is that legal users access resource on restrictions and avoid illegal access or damage system security through appropriate access control. The common access control policies contain discretionary access control (DAC), mandatory access control (MAC), and role-based access control (RBAC) [4][6] [12][13].

The National Institute of Standards and Technology (NIST) adopts RBAC as the standard, called NIST RBAC [5]. In this model, users are assigned to the appropriate role, and the access of resources is determined according to role. The RBAC model provides relevant security policy objectives, such as least privilege, static and dynamic separation of duty, and data abstraction [14].

2.3 Single Sign-On

As the Internet and related technology proliferate to support business processes, incalculable services are executed on the network at the same time. Each service provider may adopt different authentication mechanism and own individual database. Before users want to use multiple services on different service providers, users must be authenticated the identity by the different service providers at first and get exclusive username and permission to access the services.

The single sign-on model provides a single user account management interface through which all the component domains may be managed in a coordinated and synchronized manner. Users do not have to sign-on to multiple systems, necessitating an equivalent number of sign-on dialogues, each of which may involve different usernames and authentication information [1][3][16][18].

3 Single Sign-On Mechanism with RBAC

This research proposed a single sign-on mechanism with RBAC that lets users do not need to process multiple registration and login to access different cloud services on different form of cloud service providers. The proposed system architecture is shown in Figure 1, it contains the users, cloud service servers, registration information center, and RBAC database. The detailed description of each component as below:

- **Users:** The users from private cloud or public cloud. Users connected to the cloud service server as a solid line represents the user who is logged in to the registered cloud service server, and the dotted line represents unregistered cloud service servers.
- **Cloud service servers:** Cloud service providers. Cloud service servers exchange registration information with registration information center. Cloud service servers could change the user status in registration information center when users log in or log out cloud service servers, and receive the authentication information of other cloud service servers from registration information center.
- **Registration information center:** Registration information center that obtained joint certification license by cloud service servers. It will record different user's account and basic information of different cloud service servers and help the cloud service servers to authenticate unregistered users. Besides, it will connect to RBAC database to obtain the permissions of sign-on users.
- **RBAC database:** Each cloud service server established roles and permissions in this database in advance.

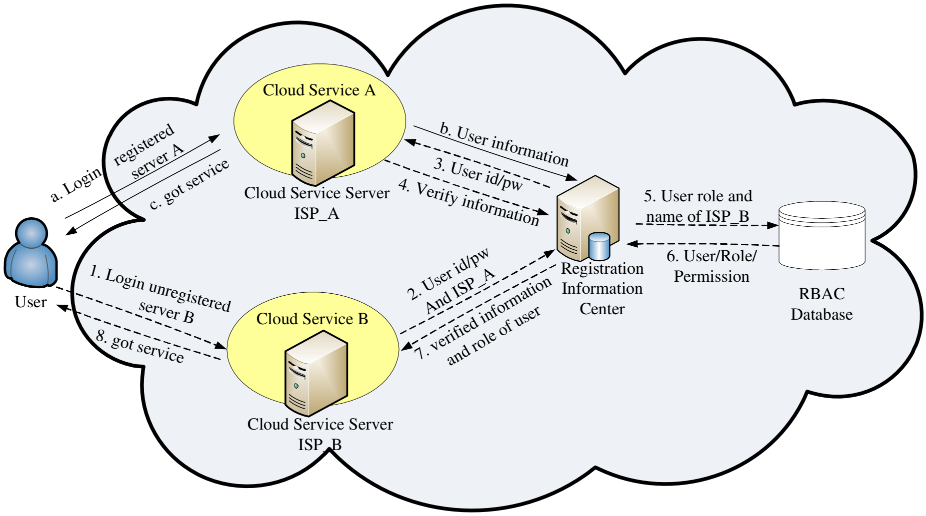


Fig. 1. System Architecture

In this study, the mechanism must satisfy the following assumptions. At first, the cloud service servers trust each other. Then the registration information center is considered as a trusted third party. Finally, using a secure channel, such as secure socket layer, i.e. SSL, to transfer information.

The user registration process for the first time is described as the following. Step a, user connected to the registration page of cloud service server A and registered related information. After the registration process, the cloud service server A will send the user information to registration information center in step b. Finally, the user will get the permission to use the service in step c. Since then user wants to sign on the registered server, only step a and c will be executed.

If user wants to log in unregistered server like step 1, the detail authentication process is described as the following. When user wants to sign on an unregistered cloud service server B (ISP_B) and provides the user account and password of registered cloud service server A (ISP_A), the ISP_B will send the information to be verified to registration information center. The information to be verified are user account (id), password (pw), and the name of ISP_A as step 2. Then, the registration information center will forward the user's id/pw to cloud service server A in step 3. In step 4, the cloud service server A received the request to verify the user identify. Cloud service server A will authenticate the information and return the result to registration information center. If registration information center receive the successful authentication, registration information center will issue the query that what are the permissions of user in cloud service server B through step 5. RBAC database retrieved the capable role and permission for a user that did not register in cloud service server B and return the result to registration information center through step 6. Registration information center the verified information from ISP_A and role/permission from RBAC database to cloud service server B in step 7. Finally, the user got the appropriate service in step 8.

4 Conclusion and Future Works

The rapid development of cloud computing, there are more and more cloud services to facilitate our lives. But each cloud service server do not exchange information with each other, users need to register multiple accounts in different cloud service servers. User not only manages several accounts, but also repeat the username and password verification in different cloud service servers. The research proposed a single sign-on mechanism with RBAC to avoid the above problem under the premise of not changing the user experiences. The proposed mechanism adopts registration information center to verify user information and forwards information to original cloud service server to authentication. In this way, the proposed mechanism achieved single sign-on and single registration. Besides, the proposed mechanism combined with RBAC database to allow users to use different cloud services according authorized permissions.

When the increasing cloud services and users, the system may be performance decreases or load imbalance. In the future, the performance of registration information center and RBAC database should be enhanced or using virtual server management to improve the effectiveness of architecture. Besides, this study hypothesized the registration information center and cloud service servers that trust each other. The verification mechanism to trust each other should be developed.

References

1. Lodha, A., Sarma, R.: A Single Sign-On Approach. Avenue a Razorfish, Inc. (March 2006)
2. Pereira, A.L.: Computing Systems Integration in Grid Computing and Cloud Computing. In: IEEE International Parallel & Distributed Processing Symposium (2011)
3. Cesnet, Web Single Sign On Systems, http://www.cesnet.cz/doc/techzpravy/2006/web-sso/ (June 1, 2012) (retrieved)
4. Ferraiolo, D.F., Cugini, J.A., Ricbard Kuhn, D.: Role Based Access Control (RBAC): Features and Motivations. In: IEEE Computer Security Applications Conference, pp. 241–248 (1995)
5. Ferraiolo, D.F., Sandhu, R., Gavrila, S., Richard Kuhn, D., Chandramouli: Proposed NIST Standard for Role-Based Access Control. ACM Transactions on Information and System Security 4(3), 224–274 (2001)
6. Ferraiolo, D., Kuhn, R.: Role-Based Access Control. In: 15th National Computer Security Conference, pp. 554–563 (1992)
7. Steiner, J.G., Neuman, C., Schiller, J.I.: Kerberos: An Authentication Service for Open Network Systems. In: Proceedings of USENIX Conference, pp. 191–200 (1988)
8. Brodkin, J.: Cloud Computing Hype Spurs Confusion, Gartner Says-Too Many Clouds-Too Little Time (2008), http://www.computerworld.com/s/article/print/9115904/Cloud_computing_hype_spurs_confusion_Gartner_says (June 1, 2012) (retrieved)
9. Mell, P., Grance, T.: The NIST Definition of Cloud Computing (2011), http://csrc.nist.gov/groups/SNS/cloud-computing/index.html (June 1, 2012) (retrieved)

10. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. Future Generation Computer Systems 25(6), 599–616 (2009)
11. Chakraborty, R., Ramireddy, S., Raghu, T.S., Raghav Rao, H.: The Information Assurance Practices of Cloud Computing Vendors. IEEE Computer Society IT Professional 12(4), 29–37 (2010)
12. Sandhu, R., Samarati, P.: Access Control: Principles and Practice. IEEE Communications Magazine, 40–48 (1994)
13. Sandhu, R., Munawer, Q.: How to do Discretionary Access Control Using Roles. In: Proceedings of the Third ACM Workshop on Role-Based Access Control, pp. 47–54 (1998)
14. Sandhu, R., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-Based Access Control Models. IEEE Computer 29(2), 38–47 (1996)
15. Subashini, S., Kavitha, V.: A Survey on Security Issues in Service Delivery Models of Cloud Computing. Journal of Network and Computer Applications 34(1), 1–11 (2011)
16. Samar, V.: Single Sign-On Using Cookies for Web Applications. In: Proceedings of the IEEE 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 158–163 (1999)
17. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A.: Cloud Computing — the Business Perspective. Decision Support Systems 51(1), 176–189 (2011)
18. The Open Group, Single Sign-On, http://www.opengroup.org/security/sso/ (June 1, 2012) (retrieved)
19. Nykänen, T.: Secure Cross-Platform Single Sign-On Solution for the World-Wide Web. Department of Computer Science and Engineering. Helsinki University of Technology (2002)

A Hybrid Multi-Criteria Decision Support Model: Combining DANP with MDS

Chen-Shu Wang[1] and Shiang-Lin Lin[2]

[1] Graduate Institute of Information Management,
National Taipei University of Technology, Taipei, Taiwan
wangcs@ntut.edu.tw
[2] Department of Management Information Systems,
National Chengchi University, Taipei, Taiwan
shiang0623@gmail.com

Abstract. Human beings encounter the problem of making decision in their daily life. However, most decision makers, when encounter the decision problems involving multi-criteria or diverse alternatives, they could not make correct decision due to lacking related decision-making information. Therefore, what is most concerned by the decision makers is how to pick out the most optimal decision-making evaluation factors and the best execution alternataive for a multi-criteria decision making (MCDM) problem. This study combined the merits of various decision-making analytic methods, namely, decision making trial and evaluation laboratory (DEMATEL), analytic network process (ANP) and Multidimensional Scaling (MDS), so as to propose a four-stage Hybrid Decision-Making Support Model (HDMSM) to assist the decision makers in making the best decision when they face a decision problem.

Keywords: Multi-Criteria Decision Making, Analytic Network Process, Decision Making Trial and Evaluation Laboratory, Multidimensional Scaling.

1 Introduction

Human beings encounter the problem of making decision all the time in their daily life. Most of these decision problems involve many factors that are to be evaluated by a decision maker, and there are usually many alternatives for one decision problem (Simon, 1977). However, most decision makers, when encounter the decision problems involving multiple evaluation factors or alternatives, often fail to make correct decision due to lacking related information (Hwang & Yoon, 1981). Therefore, what is most concerned by the decision makers is how to pick out the most optimal decision-making evaluation factors and further find out the best execution alternataive for a multi-criteria decision making problem (Yoon & Hwang, 1985).

In the field of decision science, multi-criteria decision making (MCDM) methods are most frequently used to solve the above-mentioned decision problems (Belton, 1990). The objective of MCDM methods is to help the decision makers express their preferences structure from limited number of possible alternatives, and then use

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_47, © Springer-Verlag Berlin Heidelberg 2014

various multi-criteria decision-making analytical methods, such as DEMATEL, AHP and so on, to convert the qualitative value of each evaluation criterion and alternative into quantitative weight (Buede & Maxwell, 1995), and finally, according to the priority ranking, determine the important decision-influencing factors and the ideal execution alternative (Opricovic & Tzeng, 2004). MCDM methods have been rapidly developed in the fields of management and social science in recent years, and have been widely employed in, for example, investment portfolio (Ehrgott, et al., 2004), supplier selection (Shyur & Shih, 2006), and green supply chain (Büyüközkan, 2012).

While there are numerous researches and applications of MCDM methods. However, these methods still have several disadvantages need to be improved. First, each of the MCDM methods has its own theoritical basis as well as its merits and drawbacks. As a result, when different methods are used for the same decision problem, they would usually lead to different results, so that the decision makers are at loose ends (Keeney, 1992). Therefore, it is very important that how to combine a veriety of MCDM methods in order to develop a decision support model for effectively assisting the decision makers in making a correct decision (Yang, et al., 2008). Second, the diversified social environment makes the decision problem much more complicate. In the current environment of decision problem, the evaluation factors frequently have interaction or conflict with each other, and the conventional hierarchy-based MCDM methods just could not accurately help the decision makers evaluate the factors in such decision problems (Saaty & Vincke, 1988).

In view of the above fact, ANP is one of the important MCDM methods that used in the multi-criteria decision making to effectively handle the dependence and feedback among different evaluation factors (Saaty, 2001). Before using ANP to measure the weights of the evaluation factors, a networking among all factors must first be established (Wu, 2008). However, the establishment of the networking is not included in the scope of ANP. Therefore, it need other analytical methods to find out the relation between different factors (Yang & Tzeng, 2011).

ANP can help the decision makers to find out what are the critical evaluation factors in the decision problem; DEMATEL can establish the relation between the evaluation factors; and MDS can be used in multi-alternative decision to analyze the similarity between different alternatives and further assists the decision makers to more accurately find out the most optimal alternative (Huang, 2005). The objective of this study is to combine the merits of the above three decision-making methods, so as to propose a Hybrid Decision-Making Support Model (HDMSM) to assist the decision makers in making the best decision when they face a decision problem.

2 Literature Review

2.1 Decision Making Trial and Evaluation Laboratory

Decision making trial and evaluation laboratory (DEMATEL) was originated from the Geneva of the Battelle Memorial Institute in 1973. It can effectively observe the level of mutual influence among different factors, so as to understand the complicated cause-and-effect relationship in the decision problem Fontela & Gabus (1976). The analytic process are shown as follow.

Define the Relationship among Evaluation Factors. Through literature review or brainstorming, then, listing the factors which affect the problem of decision-making, and interview the experts who are in the related field, in order to determine the relationship between each of two factors.

Establish Direct-Relation Matrix. If the decision problem with *n* evaluated factors, according to the degree of influence scores which is determined by experts, further, to establish an $n*n$ direct-relation matrix, which represent as *Z*. Among the matrix, z_{ij} represent the degree of the factor z_i effect factor z_j. The calculation is in equation (1).

$$\mathrm{Z} = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \begin{bmatrix} 0 & z_{12} & \dots & z_{1n} \\ z_{21} & 0 & \dots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & 0 \end{bmatrix} \end{array} \tag{1}$$

2.2 Analytic Network Process

ANP is a decision-making analytical method that uses network and nonlinear structure to represent a decision problem, and is developed in response to the fact that many decision problems could not be presented with the structured hierarchy. The main objective of ANP is to correct the traditional AHP, with which the problems of dependence and feedback might occur between the criteria or the layers (Saaty, 1996).

ANP mainly through Supermatrix to show the relationship and strength of graph among factors. The best advantage of using Supermatrix is it can evaluate the external and internal of dimensions dependability efficiently. In addition, we can obtain the weight of each factors through the Limit Supermatrix.

2.3 Multidimensional Scaling

Multidimensional scaling (MDS) is a data reduction method, it uses the distance or similarity between data points to locate the spatial coordinates and the relative positions of several given data in the low-dimensional space (Torgerson, 1952).

M-MDS is mainly through compute the Euclidean distance between each two factors, and show all factors in Perceptual map which has two dimensions. The similarity between two factors more stronger, the configuration of two factors more close in the map. As a result of graph can show factors more clearly and let researchers understand easily. Therefore, Through Perceptual map could show the hidden structure or spatial relation between the factors, and achieve the classification result through the spatial difference.

3 Hybrid Decision-Making Support Model

This study proposes a Hybrid Decision-Making Support Model (HDMSM). As shown in Figure 2, the decision-making procedures according to HDMSM includes total four stages, namely, Selection, Relation, Evaluation and Decision, which are described below:

Selection. From literature, proper evaluation criteria and alternatives are selected for the goal in the decision making.

Relation. To understand the relation among different evaluation criteria, it is necessary to further use the DEMATEL method to analyze the degree of mutual influence among different criteria.

Evaluation. Based on the relation among different criteria as found in the stage II, a networking structure of evaluation is plotted. Then, according to the networking structure of evaluation, an ANP expert questionnaire is designed and distributed. Further, using ANP to analyze and calculate the weights and the priority ranking of the evaluation criteria.

Decision. Use ANP to process all the evaluation criteria and the alternatives, so as to pick out the most optimal alternative and calculate the Euclidean distance among the alternatives. Then, use MDS to analyze and find out the similarity and dissimilarity among all the alternatives.

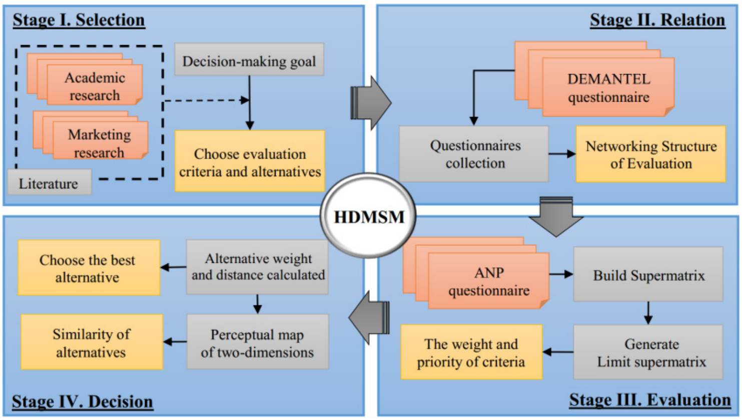


Fig. 1. Hybrid Decision-Making Support Model (HDMSM)

4 The HDMSM with Sample Demonstration

4.1 Criteria and Alternatives Selection

Before a multi-criteria decision making, it is necessary to select the criteria suitable for use as evaluation factors and to select the alternatives. The criteria can be obtained from past similar decision problems or related literature. In this study, we postulate five evaluation criteria (C1~C5) and five alternatives (A1~A5) for explaining the subsequent sample decision-making flow according to our HDMSM.

4.2 Establishment of Criteria's Relationship

After the criteria are selected, first distribute the DEMATEL expert questionnaire, so as to determine the relation and the degree of mutual influence among the criteria. The adopted scaling is the 0~3 scaling designed by Fontela & Gabus (1976); where, 0 indicates there is not any relation among the criteria, and 3 indicates there is significant relation among the criteria. Table 1 shows a sample questionnaire of DEMATEL.

Table 1. The Questionnire Sample of DEMATEL

	C1	C2	C3	C4	C5
C1	-	3	0	1	1
C2	0	-	2	2	0
C3	0	1	-	0	3
C4	2	0	0	-	2
C5	0	0	1	3	-

4.3 Plotting of Networking Structure of Evaluation

After the process using DEMATEL method, the criteria's relationship presented in the obtained results is used to plot a networking structure of evaluation, a sample of which is shown in Figure 3.

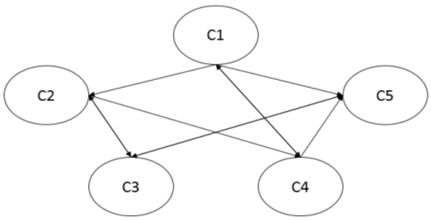


Fig. 2. The Sample of Networking Structure of Evaluation

Further, distribute the ANP expert questionnaire that is designed based on the networking structure of evaluation. For example, when use C1 as a basis of evaluation, a pairwise comparison is conducted among C2, C4 and C5. The questionnaire is rated on the Pairwise Comparison Scale proposed by Saaty (Saaty, 1980), as shown in Table 2.

Table 2. The Sample of Pairwise Comparison

Important <------------------------ Equal ------------------------> Important										
Scale	9:1	7:1	5:1	3:1	1:1	1:3	1:5	1:7	1:9	Scale
C2						V				C4
C2				V						C5
C4		V								C5

4.4 Criteria Weighting and Priority Ranking

The collected ANP expert questionnaires are calculated to acquire the eigenvectors of criteria and to form a Supermatrix. Through normalization of the Supermatrix and complex matrix multiplication, a Limit Supermatrix showing weights of the evaluation criteria can be obtained. A sample Limit Supermatrix is shown in Table 3.

Table 3. The Sample of Limit Supermatrix

	C1	C2	C3	C4	C5	Weight	Rank
C1	0.230	0.230	0.230	0.230	0.230	0.230	3
C2	0.338	0.338	0.338	0.338	0.338	0.338	1
C3	0.249	0.249	0.249	0.249	0.249	0.249	2
C4	0.033	0.033	0.033	0.033	0.033	0.033	5
C5	0.150	0.150	0.150	0.150	0.150	0.150	4
Total	1.000	1.000	1.000	1.000	1.000	1.000	-

4.5 Multidimensional Scaling and Alternative Selection

Normally, there is more than one alternative for a decision problem, for each alternative, different evaluation criteria usually have different importance levels. Therefore, after obtaining the criteria's priority ranking, it is necessary to further calculate the relative importance level of each alternative based on the evaluation criteria, so as to facilitate the subsequent alternative similarity analysis. Again, the pairwise comparison scale proposed by Saaty is used as the rating scale. Finally, the total weight and the priority ranking of each of the alternatives based on all criteria are obtained, as shown in Table 4.

Table 4. The Sample of Alternatives Priority

	C1	C2	C3	C4	C5	權重總合	Rank
A1.	0.473	0.170	0.111	0.134	0.554	1.442	1
A2.	0.059	0.055	0.423	0.095	0.102	0.733	4
A3.	0.036	0.117	0.162	0.043	0.231	0.589	5
A4.	0.149	0.396	0.271	0.480	0.064	1.361	2
A5.	0.283	0.262	0.033	0.249	0.048	0.875	3
	1.0000	1.0000	1.0000	1.0000	1.0000	-	-

Based on Table 4, the Euclidean distance between any two alternatives can be further calculated to create a Euclidean distance matrix, as shown in Table 5.

Table 5. The Sample of Euclidean Distance Matrix

	A1	A2	A3	A4	A5
A1	0.000				
A2	0.699	0.000			
A3	0.557	0.302	0.000		
A4	0.736	0.546	0.567	0.000	
A5	0.566	0.522	0.419	0.383	0.000

From the Euclidean distance matrix, we can find the coordinate positions of the alternatives in a second dimension and plot a perceptual map, as shown in Figure 4.

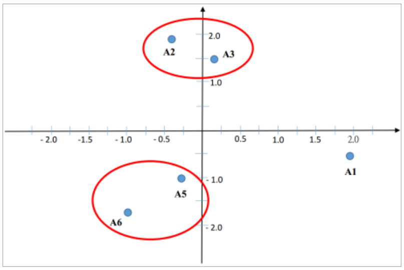


Fig. 3. The Sample of Perceptual Map

From the perceptual map, the decision maker can have a clear idea about the similarity and the dissimilarity among all the alternatives. Finally, along with the priority of the criteria and the alternatives to select the most optimal alternative.

5 Conclusions and Further Work

MCDM problem has always been a topic that can not be ignored in many different fields, such as management, social science and even engineering. The main objective of decision analysis is to help decision makers determine the evaluation criteria and find out the most optimal alternative for their decision problems.

This study combines three different MCDM methods, namely, DEMATEL, ANP and MDS, to propose a four-stage hybrid decision-making support model (HDMSM). This model can be used to effectively analyze the relational level among different evaluation criteria in a decision problem, and to find out the criteria that have most significant influence on the decision results. Then, the criteria and the decision alternatives are cross-analyzed to help the decision makers pick out the best alternative for the decision problem for execution.

The HDMSM can be applied to analyze decision problems of various issues, such as system introducing or business process reengineering. In future, HDMSM could be combined with other decision methods in order to improve the accuracy and effectiveness of this decision-making support model in handling decision problems.

References

1. Belton, V.: Multiple Criteria Decision Analysis Practically the Only Way to Choose. Strathclyde Business School (1990)
2. Huang, J.J., Tzeng, G.H., Ong, C.S.: Multidimensional Data in Multidimensional Scaling Using the Analytic Network Process. Pattern Recognition Letters 26(6), 755–767 (2005)
3. Saaty, T.L.: Analytic Hierarchy Process. Encyclopedia of Biostatistics. John Wiley & Sons, Inc. (1980)
4. Torgerson, W.S.: Multidimensional Scaling: I. Theory and Method. Pstchometrika 17(4), 401–419 (1952)
5. Yoon, K., Hwang, C.L.: Manufacturing Plant Location Analysis by Multiple Attribute Decision Making: Part I-Single-Plant Strategy. International Journal of Production Research 23(2), 345–359 (1985)
6. Simon, H.A.: The New Science of Management Decision. Englewood Cliffs (1977)
7. Opricovic, S., Tzeng, G.H.: Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS. European Journal of Operational Research 156(2), 445–455 (2004)
8. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Application. Springer (1981)
9. Buede, D.M., Maxwell, D.T.: Rank Disagreement: A Comparison of Multi-criteria Methodologies. Journal of Multi-Criteria Decision Analysis 4(1), 1–21 (1995)
10. Saaty, T.L.: Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in Complex World. RWS Publications (1988)
11. Keeney, R.L.: Value Focused Thinking. Harvard University Press (1992)
12. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publications (1996)
13. Ehrgott, M., Klamroth, K., Schwehm, C.: An MCDM Approach to Portfolio Optimization. European Journal of Operational Research 155(3), 752–770 (2004)
14. Shyur, H.J., Shih, H.S.: A Hybrid MCDM Model for Strategic Vendor Selection. Mathematical and Computer Modelling 44(7), 749–761 (2006)
15. Büyüközkan, G., Çifçi, G.: A Novel Hybrid MCDM Approach Based on Fuzzy DEMATEL, Fuzzy ANP and Fuzzy TOPSIS to Evaluate Green Suppliers. Expert Systems with Applications 39(3), 3000–3011 (2012)
16. Yang, Y.P., Shieh, H.M., Leu, J.D., Tzeng, G.H.: A Novel Hybrid MCDM Model Combined with DEMATEL and ANP with Applications. International Journal of Operations Research 5(3), 160–168 (2008)
17. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publisher (2001)
18. Wu, W.: Choosing Knowledge Management Strategies by Using a Combined ANP and DEMATEL Approach. Expert Systems with Applications 35(3), 828–835 (2008)
19. Yang, J.L., Tzeng, G.H.: An Integrated MCDM Technique Combined with DEMATEL for a Novel Cluster-Weighted with ANP Method 38(3), 1417–1424 (2011)
20. Fontela, E., Gabus, A.: Current Perceptions of the World Problematique. World Modeling-A Dialogue (1976)

Ad Hoc Social Networking on a Smart City Builder – A Practical Approach

Tang-Hsien Chang

Department of Civil Engineering, National Taiwan University, Taiwan
thchang@ntu.edu.tw

Abstract. Building a smart city has been placed on the essential projects list of various current governments. The project includes concepts such as smart home, smart market, smart health care, smart security system, smart sightseeing & leisure zones, smart power providing, and intelligent transportation systems (ITS). In fact, all subprojects rely on a seamless information or telematics system. Accordingly, build a complete seamless communication system plays an important role in smart city applications. This paper addresses available current communication technologies, particularly in the short-range communications field while long-range technologies are implemented for some services. With respect to short-range technologies, IEEE 802.11.x, IEEE 802.15.x, RFID and Barcode-like (QRcode) are the most popular protocols in use, and are applied in the service range of Near Field Informatics (N-Fi) for ad hoc social networking. This paper presents the concept architecture (networking topology and routing protocol) as well as the services provided by an ad hoc network. The case used to demonstrate the system is the Taichung bus rapid transit (BRT) system in Taiwan.

Keywords: Smart city, intelligent transportation system, near field information, MANET, VANET, real-time information, telematics, V2I, BRT.

1 Introduction

To successfully implement a smart city, selecting a proper communication technology that is inexpensive and commonly shared/owned is essential. Although 2G, 3G, and 4G protocols are in service, they are insufficient by themselves for the demands required by a smart city. We still cannot *directly* link between/among a handset with other handsets, car onboard devices, roadside units as well as indoor facilities that use dissimilar protocols. This leads to the need for an ad hoc network to fill the gap, particularly for social networking. In this paper, a Near Field Informatics (N-Fi) system is presented to compensate for the communication gap. To accomplish the N-Fi system, a VIP net is introduced herein. The letters V, I, and P denote Vehicles, Infrastructure, and Person/People, respectively. “Infrastructure” could be roadside units (such as traffic signal controllers (TSC), changeable message signs (CMS), toll gates, bus stops, convenience shops, schools, and public building facilities,). The direct linkage (through wire and/or wireless) is accomplished by matching devices using mature the “Internet of Things” (IOT) (see Fig. 1).

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_48, © Springer-Verlag Berlin Heidelberg 2014

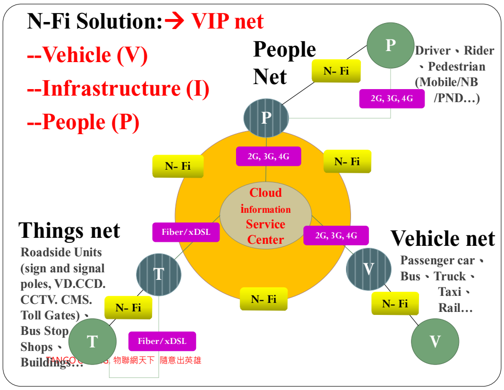


Fig. 1. The concept of VIP net

Considering direct linkage for ad hoc networking from an applications approach, the most difficult situation is providing services for moving objects (cars and/or people). Obviously, it should be implemented through wireless techniques with quick access/connect characteristics, should be easy to carry or install, and have common protocols and be inexpensive. Considering the system to concurrently support static objects and moving objects, MONET conversion to VANET is studied, where VANET is built for communication among vehicles and roadside units and drives informatics that satisfy travelers' demands (information on demand, IOD).

In this domain, Japan's VICS (Vehicle Information and Communication System) successfully applies beacon technology (2.4 GHz microwave signals for expressways and infrared for local streets) and NHK FM subcarrier signals for traffic information transmission. The USA's VII (Vehicle Infrastructure Integration) proposes a frequency of 5.9 GHz applying the IEEE 802.11p protocol for communication among vehicles (V2V) and infrastructure (V2I) [1]. Europe's SAFESPOT also includes development for V2V and V2I [2]. However, these propositions represent a barrier to international marketing due lack of portability, not sharing common frameworks, and not being easily affordable (e.g. without any additional charges to users).

This paper presents the N-Fi technology concept that is being tested in Taiwan, and specifically will be applied in the Taichung BRT system. The presentation includes the system architecture and its applications.

2 Technology Selection

Indeed, building smart cities has been placed on the essential projects list by various current governments. The project includes concepts such as smart home, smart market, smart health care, smart security system, smart sightseeing & leisure zones, smart power providing, and intelligent transportation systems (ITS). All the subprojects rely on a reliable, seamless information or telematics system. However, a complete, available communication system plays a more important role presently. The following discussion presents available communication technologies, particularly regarding short-range communications while long-range technologies are implemented using 2G, 3G, and 4Gthat have a suitable service history. Basically, current long-range communication systems are built with fixed base stations to provide their connections, called an infrastructure net, rather than an ad hoc linkage. However, Fig. 2 illustrates some short-range technologies where IEEE 802.11.x, IEEE 802.15.x, and RFID are most in use currently, and are applied in Near Field Informatics (N-Fi) services and may have potential if applied to ad hoc networking.

IEEE 802.11x is widely used for electronic devices such as WiFi devices (IEEE 802.11a,b,g,n). Almost all personal handsets are equipped with WiFi. Many cities in the world offer free service in public sites such as airports, public transportation stations, and social activity squares. Then, one can easily access the Internet or other network to obtain information or perform their business online. However, it is difficult to obtain WiFi service when one is moving above 20~30 kilometers per hour (kph). The IEEE 802.15.x protocol generally equipped inside personal handsets is Bluetooth (IEEE 802.15.1), and an alternative is Zigbee (IEEE 802.15.4). Bluetooth has a 1~2 Mbps transfer rate and Zigbee has only 250 Kbps. Nevertheless, Zigbee allows connection in high speed moving vehicles due to Zigbee's short linking attribute. Passive RFID is also currently popular for very short-range requirements, approximately 10 cm or less, such as for contactless tickets or door-keys. Extending passive RFID's reader power allows detection of the tags mounted on high speed vehicles, where the detection range can reach 10~15 meters. Active RFID is utilized for ranges of 10-100 meters that are accomplished using Zigbee (see Table 1).

In conclusion, WiFi and Bluetooth will be the potential technologies for ad hoc social networking for large message packets delivery if some barriers are relieved, and a protocol such as Zigbee is able to deliver short message packets. Therefore, a tightly (and perhaps even loosely) coupled network using Zigbee, WiFi, and/or Bluetooth has its potential in today's markets.

3 N-Fi System Architecture

N-Fi is the acronym for the Near Field Informatics domain that utilizes the combined chips of Zigbee/WiFi/Bluetooth as the ad hoc basis for VIP linkage, particularly for social networking. They all operate at 2.4 GHz. Fig. 3 illustrates the N-Fi topology for the proposed implementation. It connects four layers including Vehicles (Occupants) and/or people/persons, Roadside Units (RSU), Operational Center, and Web Users. The structure supports and is consistent with the VANET framework.

Basically equipped with a N-Fi module (using Zigbee, WiFi and/or Bluetooth) in vehicles' onboard units or mounted to handsets (mobile phones), the Zigbee mode will drive short messages and the WiFi/Bluetooth mode will drive transferring a file or picture. The protocol that drives an application depends on usage situations. In the example application case of ITS, the equipped RSU receives all nearby OBUs and/or Handsets' identification (ID) and then sends its own ID and information related to a vehicle's passing time and local speed to the operational center. The operational center stores and processes all the messages (data). The center may issue predicted traffic information through XML to web pages and OBUs via RSU. In N-Fi, the applied protocol is guided by and follows the NTCIP (National Transportation Communication for ITS Protocol). Due to the different networks and their different maximum transmission units (MTU), all packets will be regrouped before transmitting.

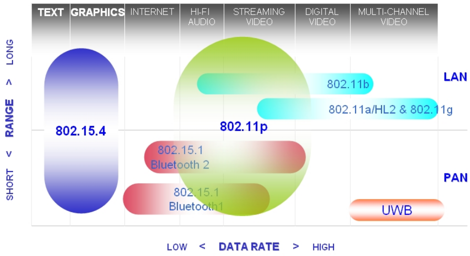


Fig. 2. Short Range Communication technology Profile

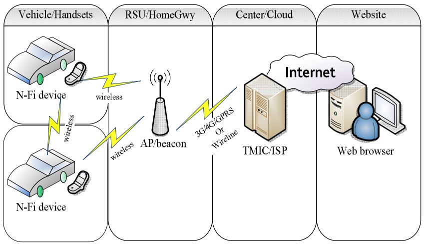


Fig. 3. N-Fi simplified Topology

Table 1. Short Range Comm. Tech and Attributes

Comm Tech	Blue-tooth	ZigBee	Wi-Fi	RFID
Moving constraint	<5kph	50~100 kph	<30kph	30~80 kph
Service range	10~100 m	10~1000 m	10~100 m	< 10m
Connection speed	5~10s	<1s	1~10s	<1s
Bandwith	~1Mbps	250Kbps	>1Mbps	-
Transmission	two ways	Two ways	Two ways	One way
Handset	Build-in	Extra	Build-in	Extra (eTag-like)

To avoid errors occurring when regrouping, a formatting protocol for transmitting packets is applied, as shown in Fig. 4. Except for the “Information Field”, deleting some redundant TYPE-OCTETs in the “Header Field” in N-Fi is considered. To integrate “SEQ”, the OID type packets defined in “OBJECT ID” are deleted and replaced with “Util” for judging OID. In processing for ad hoc transmission, each message has its sequence number (SEQ) and OID for broadcast (even in unicast and/or pancast scenarios), such that each node processes each message once only. If the same SEQ message arrives a second time, the node will delete the duplicate message. The SEQ is assigned by the message source node, and cannot be changed by any other node. Since each node (vehicle or person) does not know other neighbored nodes' positions or locations, broadcasting with different power levels over a short interval determines the distance ranges for neighbor nodes (via RSSI or received/not received criteria). Then, the transferring nodes can be set to communicate with further nodes based on power levels to avoid collision and broadcast storms caused by using the same power levels for all transmissions. In general, message delivery should be set so it does not exceed a transfer limit and/or a scheduled time limit or active/expire time to limit the number of transfers. The scenarios described above activate between/among mobile phones or similar devices that are equipped with N-Fi modules.

1	1	2	1	6	1	1			1	1	1
H1	H2	Length	SEQ	Device Address	Util	OidL	Object ID	Information Field	H1	Hend	CKS

Fig. 4. Sample of the N-Fi protocol format

4 Application Case: BRT in Taichung

Urban traffic often perplexes people while traveling. Many dense cities have constructed Mass Rapid Transit (MRT) to provide a core transportation system. However, such a system is expensive, no matter what the construction and operational costs. Bus Rapid Transit (BRT) as a complement and/or improvement to an existing transportation system is relatively mature in many cities around the world. In general, BRT operates in an exclusive lane, but does not have exclusive right of way. BRT improves the efficiency of an existing urban transportation system by limiting the financial investment and construction period compared to MRT. The key technical difference that distinguishes BRT from traditional bus operations is the emphasis on short-range communication applied between the BRT buses and roadside signal controllers. Therefore, a valid designed logic and rules for traffic control are very important when designing a BRT.

In Taichung, the key components of BRT include (1) onboard devices, (2) applied communication technologies, (3) station platforms, (4) roadside traffic signal controllers at intersections, (5) the BRT operational center, and (6) the traffic control center. Zigbee is selected as the short-range communication mode.

An onboard device must have (1) a computer driving the functions for GPS location positioning, location-based information broadcast by an LED board for passengers, computing available/suggesting speed for approaching and passing through an intersection displayed on the dashboard, positioning for entering a station's corresponding platform and exact gates, driving door and gate opening or closing synchronously, and performing car driving data recording; (2) a communication module with a 3G channel for long-range mobile communication with the operational center, and a IEEE802.15.4 device for short-range communication with traffic signal controllers and station gate controllers; (3) a controller for executing traffic signal preemption commands when appropriate.

The applied communication technologies include: (1) a 3G mobile module for communication with the operational centre that transmits a vehicle's location and clock time, emergent vehicle mechanical/physical conditions, on-board messages, operational centre commands to the driver or messages to passengers; (2) an IEEE802.15.4 (Zigbee) module for communication with the approaching traffic signal controller that transmits the vehicle's location to the controller in meters, the vehicle ID, signal timing phase and its remaining time limit in seconds, and the vehicle's authorized level; (3) an IR (infrared) device for communication with platform gate controllers to position bus doors exactly and to open station gates sequentially.

The traffic signal controller has an important role in the BRT system. It measures the oncoming BRT vehicle's location, computes the current traffic control phase and timing distribution, authorizes whether the BRT vehicle is granted right of way or not, controls traffic lights (on/off), etc. Therefore, the equipped devices include a CPU processor board, a manual and display board, a connector for vehicle detectors, a communication processor with access to a fiber net to the traffic control center, an IEEE802.15.4 module to provide vehicle onboard connection, etc.

The operational centre handles vehicle scheduling, position/speed control, information broadcasting, emergent processing, station surveillance and control, ticketing clearance, staff management, vehicle and device maintenance system, line and route plans, etc. The traffic control centre handles integration of the controllers and works to optimize the entire traffic network, monitors BRT preemption performance and its influence on other modes, and authorizes intersection signal controllers for BRT right of way preemption when appropriate.

Based on the vehicles, traffic controllers along the BRT line, and stations all equipped with Zigbee modules, the fundamental ad hoc network is formed according to the N-Fi architecture. The next stage is to provide a SD card-based handset Zigbee module or a Bluetooth channel link to mobile phones. A test has been conducted and resulted in perfect operation. On the other hand, a loosely/tightly coupled module for WiFi and Zigbee (relieved barriers version) is also in progress. We foresee that ad hoc networking as well as ad hoc social networking platforms will be implemented in the Taichung BRT system, and will achieve the target network shown in Fig. 1.

5 Conclusion

This paper describes how to build a component of a smart city by means of developing an ad hoc networking system. The important issue is to select a communication technology suitable for ad hoc networking that requires compatibility with popularly owned devices, particularly compatible with communication protocols already built-in to mobile phones. WiFi and Bluetooth are feasible for this purpose. However, Zigbee is also an option because it provides easy and affordable access when equipped on high speed vehicles. Currently, it is relatively easy to connect a Zigbee SD-card module in series with a Bluetooth module within a mobile phone.

The final N-Fi application for the Taichung BRT system will consist of Zigbee and WiFi to provide the ad hoc network to fulfill requirements for the VIP net.

References

[1] Farradyne, P.B.: Vehicle Infrastructure Integration(VII):VII Architecture and Functional Requirements. ITS Joint Program Office. US Department of Transportation, Washington, D.C. (2005)
[2] Brignolo, R.: Co-operative Road Safety - The SAFESPOT Integrated Project, Vienna (May 2006)
[3] http://www.ntcip.org

A Message Broadcast Model for Train-to-Train Communication Network

Pengyu Liu[1], Bo Ai[1], Yan Li[1], and Ruoyu Sun[2]

[1] State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University
Beijing, 100044, P.R. China
[2] Department of Electrical Engineering, University of South Carolina
Columbia, SC, 29208, USA

Abstract. The railway safety requires that the communication network provides a sufficiently reliable transmission of control message. Though the Communication Based Train Control (CBTC) system is applied, there are still some railway accidents. This paper concentrates on a novel train-to-train communication network, proposes the scheme of message-broadcasting in the scenarios of the regional networks and shunting yard. In the regional networks, the train transmits motion state information to the neighboring and gauge-changing trains while it is in operation. In the case of an emergency, it broadcasts the pre-warning messages to the backward trains. In the shunting yard, each railway carriage of the freight train is equipped with an antenna and a reduction gear. When the railway carriages operate on the same track, they broadcast shunting yard messages to each other to keep distance for safety. Once the speed of the railway carriage surpasses the limit, the reduction gear should begin to work. The train-to-train communication network reduces the cost of the maintenance of the infrastructure, avoids the overall malfunction from the control center system and decreases the time delay produced by the wired-wireless system. In comparison with the wired-wireless system, simulation results show that the train-to-train communication network with the scheme of message-broadcasting improves the network performance to ensure fast, reliable and efficient transmission of the control message.

Keywords: Train-to-train communication network, Message propagation model, Classification of the pre-warning messages.

1 Introduction

In the past few years, train-to-train (T2T) communication network has increased its popularity when applied to the safety of railway operation. Nowadays the mainstream technique for railway safety is the CBTC system based on the GSM-R which acts as a radio interface to link the trains with the control center system, in order to ensure that the trains are monitored by the real-time system and are performing within a certain safe distance with each other [1]. Despite the fact that the CBTC system has provided technique of accurate positioning and a rapid exchange of the motion state and control

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_49, © Springer-Verlag Berlin Heidelberg 2014

message, there exist a number of accidents to exert a sever threat to the passengers' personal safety [2]. This safety system still has some shortcomings:

1. The infrastructures along the wayside are inflexible and expensive to maintain [3]. The infrastructure is the only tie to associate the trains with the control center system. It costs great expense to overhaul and make them work well, once they are damaged and not fixed in time, an accident would occur to result in a huge harm to the safety of people's lives and property. And it is inflexible for the safety components along the wayside to detect the positions of the train, for the simple reason they are fixed and uneasy to move.

2. The malfunction of the control center system may lead to an overall safety accident all over the railway network [3]. If a mistake has been made by the control center system, the trains operating on the rails might suffer from a terrible collision with each other or with vehicles on a level crossing.

3. The time delay results from the master-slave control mechanism between the CBTC and trains. The existing train control system exchanges messages with the trains by GSM-R. From transmitting practical operating data to receiving the control messages of the control center system are there some time delays for the trains, which may be harmful to the railway safety, due to the fact that the high speed trains are sensitive to the delay which may result in a quite long operating distance.

Based on three reasons mentioned above, it is dispensable to develop a novel communication network to reduce the outlay on the maintenance, to avoid the overall malfunction from the control center system, and to decrease the time delay of the transmission of the railway control message. On the basis of the thought of the ad-hoc inter-vehicle communication, it is train-to-train that is able to achieve the target of novel communication network, which has been studied by German Aerospace Center (DLR), in recent years, to design an on-board equipment called Railway Collision Avoidance System (RCAS) [4]. Using the global satellite navigation system GALILEO, this system determines and broadcasts information about position, movement vector to other trains around in its coverage for collision detection. When a collision is detected, the train transmits pre-warning messages to others to avoid the accident [5]. Though the RCAS has made some progress in the channel modeling, Medium Access Control (MAC) layer scheme and basic RCAS message format [3-6], there are still some aspects ignored in the classification and broadcasting scheme of the message in the scenarios of regional networks and shunting yard.

The paper is structured as follows: Section 2 designs the detailed scheme of the message broadcasting in the regional networks, including the supplement of the classification and format of the pre-warning messages, while in Section 3 we propose the message broadcasting scheme in the shunting yard, dealing with the classification and format of the shunting yard message. Section 4 describes the main parameters of simulation using NS2.33 with simulation results shown. We compare the performance of T2T network with the wired-wireless network in terms of packet delay, packet loss rate and throughput. Finally, conclusions are drawn in Section 5.

2 Proposed Message Broadcast Model in the Regional Network

This paper provides a feasible model of the message propagation which depends on the thought of the ad-hoc inter-vehicle communication. This model is divided into two cases including the normal operation and emergency. When in its operation, the train exchanges the movement information with neighboring trains to detect the potential collision, and it broadcasts the pre-warning messages in the case of the emergency.

The trains are assumed to be equipped with the location detection system to detect the positions applying the COMPASS Navigation Satellite System (CNSS). The research is rooted in the background of the high-speed rail the maximum speed of which is supposed to be 360 Km/h. The tight time interval between trains spaced by automatic block signals is in the assumption of 10 Km and there is a direct wave [7].

Suppose the train we are observing is a reference train. The reference train disseminates the motion state messages to the neighboring trains on the same track and to the gauge-changing trains on the sideward track. The movement state message is composed of the track id, train id, position, direction, velocity, forward distance, backward distance and braking distance [5]. Gauge-changing id, de/acceleration, and station mode are intended to be supplement to this message. As for the position, using CNSS to acquire the location in the form of x and y coordinates format, the position information of the latitude and longitude of the reference train is known by itself which is displayed on the on-board electronic map. The map is able to convert the geographical information, x and y coordinates, to the terrain information. The corresponding information of geographical coordinates and terrain is stored in the electronic map in advance, and the driver can keep updated knowledge of terrain condition while the train is marching [5].

The neighboring and gauge-changing trains are in close touch with the reference train. The neighboring trains are occupying the same track with the reference train in the tight safe following operation distance, while the gauge-changing trains on the sideward track would move to the track of which the reference train is in occupation. The train receives the information and then is identified by the gauge-changing id in the movement state message transmitted by the reference train. Other rains on the identical track with the track-changing train will not deal with the information. Not only does this method ensure the well-targeted transmission of the useful messages, but also largely decreases the quantity of the messages to handle.

When the motion state messages are well received, the neighboring and gauge-changing trains process these data, associating the provided messages with their own states of movement. They can detect the possible collision and danger. On the same track, the forward and backward trains adjust their distances from the reference one, taking information about position, velocity, distance and de/acceleration into account, to avoid the accident. With regard to the sideward track, the track-changing trains constantly counts the time to arrive at the gauge-changing position, and compares it with the receiving information from the reference one to avoid the crash.

Especially the trains can also monitor the occurrence the geological hazards around, such as debris blow, landslide, avalanche, extreme weather and broken rail using the sensors [3]. The state of motion message is in fixed length and data rate is variable depending on the speed and neighboring traffic situation [5]. The 141-bit length message begins with 4 bits which are describing the type of the message, the next 30 bits are in the meaning of the track id, and the following 1 bit indicates the appearance of the gauge-changing. The subsequent part successively is message version, train ID, station mode, direction, latitude, longitude, velocity, acceleration, forward distance, backward distance, braking distance, pre-warning type and acknowledgement [5]. We propose the part of gauge-changing id, station mode, de/acceleration, pre-warning type and acknowledgement. The message format is shown in Fig.1.

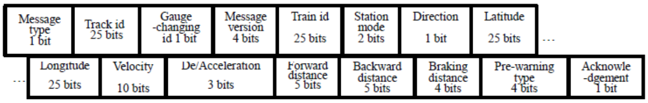


Fig. 1. Movement state message format

Through the first block of the message, the train judges the type of the message, if it is the motion state message, the train then inspects the track id determining where the train is located, if it is not on the same track, the gauge-changing will be checked to decide whether to discard the message or not. Message version is applied to ensure the message is consistently new. In the station mode, the train entering into the station keeps in touch with other trains staying in the station to prevent from the impact. The last two parts are intended for the pre-warning information, such as message type, next plan and acknowledgement.

Table 1. Classification and Description of The Pre-Warning Message

Type	**Description**
Rail collision warning	Warn other trains of the possible collision
Forward barricade warning	Warn other trains of the forward barricade
Out-of-gauge load warning	Warn other trains of the out-of-gauge load
Rail change warning	Warn other trains of the gauge-changing
Broken rail warning	Warn other trains of the broken rail
Accident in compartment warning	Warn other trains of the occurrence of the accident in the compartment
Dangerous good warning	Warn other trains of the danger of good
Extreme weather	Warn other trains of the extreme weather, (such as dense fog, heavy rain and blizzard)
Geological hazard warning	Warn other trains of extreme weather, (such as debris blow, landslides and avalanche)

After the collision is detected, the reference train disseminates the pre-warning messages to warn other trains of the potential danger. Based on the information contained in the pre-warning messages, other trains, on the same track, adjust their velocities and accelerations to keep a safe distance from each other or to take feasible measures if the accident is severe. For the gauge-changing trains on the sideward track, they might base their routes on the emergency to evade being in a dilemma. Because the railway line is extending a hundred kilometers and the propagation of the radio waves is constrained by the distance, in the process of the message broadcast, each train receiving the message should act as a relay to expand the range of dissemination. In order to ensure the reliable transmission, an acknowledgment is required. According to the kind of the danger, such as collision, forward barricade and out-of-gauge load, the pre-warning messages can be classified into various types, such as rail collision warning, forward barricade warning, out-of-gauge load warning, dangerous good, warning broken rail warning and dangerous good warning [3], we propose rail change warning, extreme weather warning, accident in compartment warning and geological hazard warning as supplement. The detailed classification and description of the pre-warning message is listed as follows. The format of the pre-warning message is basically similar to the movement state message format, but has differences in the message type, pre-warning type and acknowledgement. In the movement state message, the message type indicates that the type is the state of motion, and at this time the pre-warning type and acknowledgement are all set to 0. While in the type of pre-warning, according to the type of the danger, they are set to corresponding values.

3 Proposed Message Broadcast Model in the Shunting Yard

The shunting yard is used to process the register of arrival, disorganization and marshaling of large quantity of trains. When the train enters into the shunting yard, it would broadcast the Hello Packets to the trains around to inform its incoming and to transmit the shunting operation plans included in the shunting yard message to conduct the marshaling. In the shunting yard message is there a carriage ID by which the marshaling of the carriages of different trains are identified and organized. The engine will also receive the Hello Packet and shunting yard message which takes the responsibility for drugging theses carriages to the hump to enter into the track.

Each carriage of the freight train requires an antenna and a reduction gear. On the basis of the shunting operation plan, the shunting yard messages make the railway carriages belonging to different freight trains communicate with each other. They propagate the shunting yard messages among them to exchange the motion state and to prevent from the collision, while in operation. If the speed of the railway carriage surpasses the limit, the pre-warning messages are broadcasted and reduction gear should begin to function. The shunting yard message is 54-bit length, the function of each block is similar to the motion state message, and particularly carriage id is applied to mark different carriages together to form a self-organized network. An acknowledgement is also needed to make sure that the pre-warning message is well received. The format of the shunting yard message is shown below in Fig.2.

Message type 1 bit	Message version 4 bits	Carriage id 25 bits	Velocity 10 bits	De/Acceleration 3 bits	Forward distance 5 bits	Backward distance 5 bits	Acknowle -dgement 1 bit

Fig. 2. Shunting yard message format

4 Numerical Analysis

4.1 Simulation Environment

This paper proposes an efficient scheme which introduces an idea of the propagation model of motion state and pre-warning message in the scenarios of the regional network and shunting yard in the T2T communication network. The simulation environment of the proposed model is NS 2.33 that is an open source simulator and concentrates on the network simulation [8]. Applying this simulator, a train considered as a moving node could be simulated to set its operating state including initial position, speed and moving area, which is rewarding in the collision-avoided applications.

The simulation scenarios consist of both the T2T network and wired-wireless network, especially the T2T network is composed of two moving nodes to perform direct communication to transmit the motion state and pre-warning messages. There are four nodes in the wired-wireless network including two wireless nodes, one base station and one wired node. The wired-wireless network is used to simulate practical railway communication network, base station is acting as a base station in GSM-R to become a gateway between wireless and wired network, and the wired node is performed as the control center system. The wireless nodes communicate with the wired node via the base station to acquire the control messages instead of the direct communication between them. Because of the limitation of simulation software, the channel model applies two-ray model, and the MAC layer adopts the 802.11 protocol, as a result the distances among nodes are within 1000m and the moving speed is around 20m/s. The detailed simulation parameters for the T2T and wired-wireless network are shown below.

4.2 Performance Evaluation

The model performance is measured in terms of packet delay, packet loss rate and throughput by altering the data rate of transmission message. The results indicates that T2T network worked better using the train-to-train direct communication compared to wired-wireless mixed communication in the regional network. The comparison shows an improvement of network performance index ensuring the fast, reliable and effect transmission of the control message.

Table 2. Simulation Parameters for T2T Network and Wired-Wireless Network

	T2T network	Wired-wireless network
Simulation area	(600, 2)	(600,300)
Initial position of moving nodes	Moving node 1: (0,0) Moving node 2: (200,0)	Moving node 1: (0,0) Moving node 2: (200,0)
Position of base station	\	(100, 100)
Position of wired node	\	(100, 300)
Maximum speed	20m/s	20m/s

Fig. 3 shows an obvious difference of packet delay between the T2T communication network and wired-wireless network in which a wired node and a base station are existing. There are two moving nodes, at 0s, one is in the position of (0, 0), while the other is in the position of (200, 0). For the wired-wireless network, the base station is located at (100, 100) and the wired node was at (100, 300).The moving nodes considered as the trains are operating at the maximum speed of 20 m/s, and the simulation time is 20s, as a result the moving field in the direction of X axis was 0-600. Packet delay is the time difference between sending and receiving time. In Fig.4, the packet delay in the T2T network is staying at a low level, while it begins to increase rapidly at data rate of 340kbits/s, and from 480kbits/s, it stops the rapid growth to fluctuate in a small range. The packet delay in the T2T network, in contrast to the rapid increase of that in the wired-wireless network, varies in the area of lower value, which ensures the real-time transmission of the control message to make the train take measures to face an emergency.

Packet loss is a rate to measure the efficiency of the receiving end. Fig. 4 shows some similarities of the variety of the packet loss rate between wired-wireless and T2T network. They initially are staying at a fixed value, 0 for wired-wireless network and 0.7% for T2T network. Therefore, for the low data rate, the performance of the wired-wireless network is superior to that of the T2T network. But once the data rate arrives at 260 kbits/s, the packet loss of the wired-wireless network starts to grow in acceleration. At data rate of 460 kbits/s, the packet loss in the wired-wireless network reaches at 41.8%, while it remains 0.7% for the T2T network, and it begins to increase rapidly at 480kbits/s. The lower packet loss rate of the T2T network at higher data rate, compared to the wired-wireless network, suggests that T2T communication network guarantees the fast and safe transmission of the message.

Fig. 5 shows the variation of the throughput for the wired-wireless and T2T network. In the wired-wireless network, it does not stop the rapid increasing trend until the data rate approaches 280kbits/s and the maximum throughput is 266.78kbits/s with the data rate of 500kbits/s. From 280kbits/s, the slow change of the throughput restricts the elevation of the data rate. For the T2T network, it has the same increasing trend as that of the wired-wireless network, but the velocity of increase of the throughput in the T2T network is greater. When the data rate reaches 460kbits/s, the throughput remains almost constant around 921kbits/s. In contrast to the wired-wireless network, T2T network has higher throughput and data rate to hold a large scale of message exchanges.

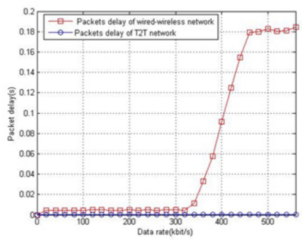


Fig. 3. Packet delay of wired-wireless and T2T network

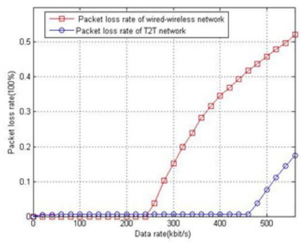


Fig. 4. Packet loss of wired-wireless and T2T network

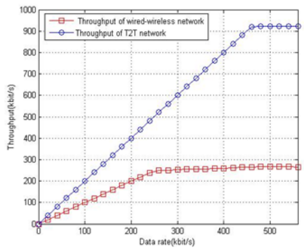


Fig. 5. Throughput of wired-wireless and T2T network

5 Conclusion

A message propagation model for T2T communication is proposed on the basis of velocity and distance parameters of the moving nodes. The neighboring and gauge-changing trains are receiving the motion state information transmitted by the train in operation. When an emergency happens, the pre-warning messages are broadcasted to the backward trains to avoid the potential accident. The model was in comparison with that of the wired-wireless network and provided effective results in terms of performance index, such as packet delay, packet loss rate and throughput in the regional network. The simulation is performed by using the propagation of the pre-warning message in the T2T and wired-wireless network. The simulation results show T2T network improved the network performance making sure that the control messages are transmitted fast, reliably and efficiently.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant 61222105, the Fundamental Research Funds for the Central Universities under Grant 2010JBZ008, Beijing Municipal Natural Science Foundation under Grant 4112048, project of State Key Lab under Grant No. RCS2012ZT013, key Project for Railway Ministry of China under Grant 2012X008-A, interdiscipline cooperation projects of the New-Star of Science and Technology supported by Beijing Metropolis under grant no. xxhz201201, and the Key grant Project of Chinese Ministry of Education under Grant 313006.

References

[1] Gao, J.J., Chen, X., Tan, Z.H.: The CBTC Technique in the ITS of high speed rail. Electric Railway 6, 41–44 (2006)
[2] The list of severe rail accident in China from 1970, http://www.xcar.com.cn/bbs/viewthread.php?tid=7481000
[3] Strang, T., Zu Hörste, M.M., Gu, X.G.: A railway collision avoidance system exploiting Ad-hoc inter-vehicle communication and GALILEO, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.4267
[4] Garcia, C.R., Lehner, A., Strang, T., et al.: Channel Model for Train to Train Communication using the 400 MHz Band. In: IEEE Vehicular Technology Conference, pp. 3082–3086 (2008)
[5] From the skies to the rails: DLR scientists unveil a new collision avoidance system for trains, http://www.dlr.de/en/desktopdefault.aspx/tabid-3086/4804_read-24385/
[6] Lehner, A., Rico-Garcia, C., Wige, E., Stang, T.: A Multi-Broadcast Communication System for High Dynamic Vehicular Ad-Hoc Networks. In: International Conference on ICUMT 2009, pp. 1–6 (2009)
[7] Garcia, C.R., Lehner, A., Strang, T., Röckl, M.: Comparison of Collision Avoidance Systems and Applicability to Rail Transport. In: 7th International Conference on ITS, ITST 2007, pp. 1–6 (2007)
[8] Zhao, P.: The operation and reorganization of the high speed train. China Railway Publishing House (2009)
[9] The Network Simulator: ns-2, http://www.isi.edu/nsnam/ns/

Urban Viaduct Channel Characterization of Train-to-Train Communication at 900 MHz

Pengyu Liu[1], Bo Ai[1], Yan Li[1], and Ruoyu Sun[2]

[1] State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University
Beijing, 100044, P.R. China
[2] Department of Electrical Engineering, University of South Carolina
Columbia, SC, 29208, USA

Abstract. Railway safety has seen attentions from tremendous researchers. Train-to-train (T2T) communication has been studied as an assisting method to enhance railway transportation safety. The key characteristic of train-to-train communication is that direct communication among trains is conducted without help of a base station. This requires a thorough investigation of T2T propagation channel characterization. This paper presents a geometry-based stochastic modeling (GBSCM) model for T2T communication in urban viaduct environment. The proposed model is used to explore T2T urban viaduct channel characteristics, i.e., path loss, root-mean-square delay spread (RMS-DS), and stemming from these a tapped-delay line (TDL) model is developed. The V2V channel's non-stationarity leads to time variation of the Ricean K-factor, thus we investigate this as well as the large/small -scale envelope covariance. The GBSCM model is valid for effectively predicting T2T urban viaduct channel characteristics.

Keywords: Path loss, RMS-DS, Tapped-delay line model, T2T.

1 Introduction

Railway is an effective and economical means to connect places at thousands of kilometers to transport passengers and goods. Railways are viewed as a crucial means of current transportation systems. Compared to other transportation methods, it is the advantages that railways both have convenient transportation and low expense. It can be seen from the real world that terrible accidents will always occur in the railway transportation. Our mainstream technique for railway transportation is the Chinese Train Control System Level 3 (CTCS-3) which uses the Mobile Communication for Railways (GSM-R) as a wireless radio interface to link trains. From the statistics provided by the American Federal Railroad Administration (FRA) in the United States, there were about 8221 accidents occurring in the past four years [1]. The reason is that the traffic condition around a train is informed by an operation control center. Once the operation center does not broadcast control messages, an accident will unavoidably happen. So we should develop a novel technique as a subsidiary method to make the train operation safer and more accurate. This technique makes the

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_50, © Springer-Verlag Berlin Heidelberg 2014

train conductors to keep update with accurate information of traffic conditions in their vicinity [2].

The train-to-train (T2T) communication, based on inter-train multi-hop communication, can be applied to detect a potential collision and then broadcast pre-warning messages to other trains on the same and neighboring tracks. When potential accidents are detected, the train-to-train communication will immediately broadcast messages to other trains in the vicinity with several solutions provided to the driver. Furthermore, its advantages also including reducing outlays on infrastructure maintenance for base stations [3]. In prior work, reference [2] discussed the RCAS approach consisting only of mobile ad-hoc components without the necessity of extensions of the railway infrastructure. The author of [3] described an overview of the state of the art in collision avoidance related with transportation systems for maritime transportation, aircraft, and road transportation, and the RCAS is introduced. Reference [4] proposed a channel model for direct T2T communication appropriate for the 400 MHz band, and reference [5] presented an infrastructure-less cross-layer train-to-train communication system exploiting all characteristics of a pervasive computing system, like direct communication in mobile ad-hoc networks. The author of [6] conducted a detailed surveillance strategy concept for a Rail Collision Avoidance System (RCAS) that is based on direct T2T communication. Reference [7] designed an infrastructure-less ad-hoc inter-vehicle communication system that fulfills these requirements with respect to the boundary conditions in the railway environment. Reference [8] presented analysis and results of a comprehensive measurement campaign investigating the propagation channel in case of direct communication between railway vehicles. The author of [10]-[11] proposed a novel physical–layer communication mechanism and model in high speed rail. The mechanism of this model lies in the idea that a source train uses trains on other tracks as relays to transmit signals to destination train on the same track, and they evaluated the outage probability of the novel communication model in high speed rail.

The geometry-based stochastic modeling (GBSCM) approach was first found in channel modeling for cellular communication [12]. The author of [12] provided a generic model for a multiple-input/multiple-output (MIMO) wireless propagation channel in cellular macro- and microcells. Then the authors of reference [13] further extended it to vehicle-to-vehcle (V2V) communication, presenting a V2V GBSCM model for highway and rural scenarios. In their model, they defined three different scatterers (mobile, static and diffuse) with unique distance-decay statistical fading properties. So far, there is no T2T channel model for urban viaduct channel at 900 MHz. In urban environment, rich scatterers along the rail track can provide multiple reflected-paths added at the receiver (Rx) to reduce the path loss. To fill this gap, this paper presents T2T GBSCM model in the urban viaduct environment. We characterize the path loss, root-mean-square delay spread (RMS-DS), and tapped-delay line (TDL) model. We also investigate the time-variant Ricean *K* factor and the large/small -scale envelope covariance.

The rest of the paper is organized as follows: In Section 2 we provide GBSCM modeling results. Section 3 concludes this paper.

2 GBSCM

2.1 Setting and Model Parameters

The propagation environment of the GBSCM is simulated by using multiple types of randomly distributed scatterers. These can be mobile scatterers, static scatterers, and diffuse scatterers, which provide multiple reflected paths to the receiver. The distance-dependent amplitudes of all reflected multipath components can be computed by geometrically tracing propagating rays from Tx to Rx (via single bounce), and these are summed up to compute the received amplitude at the receiver. The mobile scatterers are moving trains on the track; static scatterers are trees and buildings; diffuse scatterers model the remaining effects not modeled by the other specular reflections.

Fig. 1 shows the geometry of the scattering environment for modeling. There are two trains 3000 m away from each other on the neighboring track at beginning in the simulation. They move towards each other at a constant velocity of 100 m/s. We vary the time at every 0.005 s with total simulation time 30 s. The “dots” distributed on both road sides correspond to static or diffuse scatterers. They are generated by uniform distributions (*x*-coordinate) and normal distributions (*y*-coordinate), respectively. The heights of the scatterers are uniformly distributed (0~20 m). Multiple rays are reflected from the mobile, static and diffuse scatterers and the phases are traced according to distance yielding constructive and destructive combining at the receiver. All the related parameters for T2T GBSCM model in the urban viaduct setting are presented in Table 1.

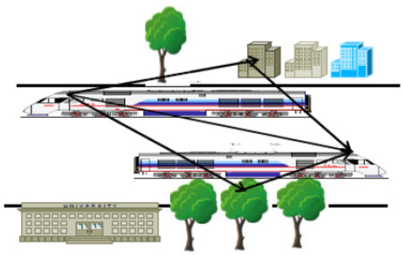


Fig. 1. Geometry of scattering environment for T2T urban viaduct channel

Table 1. Simulated Parameters for T2T Urban Viaduct GBSCM (Notation U denotes a uniform distribution on (0,1))

Simulation range	X_s (m)	200
	X_{max} (m)	3000
Tx&Rx	H_1/H_2 (m)	4.3 m
	X_Tx (m)	0
	X_Rx (m)	[0,3000]
Scatter Density	Den_s	0.1
	Den_d	0.04
Path loss Exponent	n_{LoS}	2
	n_{DI}	2.5
	n_{SD}	2U+2.8
Reference Gain	G_{LoS}	-56
	G_{DI}	$-110+13n_{SD}$
	G_{SD}	-53

We have generated specific parameters from the measurement settings for our urban viaduct and Tx/Rx characteristics. Parameters X_{max} is the main parameters for the urban viaduct, denoting the maximum simulated distance. Parameters *X_Tx* and *X_Rx* are the x-coordinates of the Tx and Rx. Parameter X_s defines the size of the region within which static and diffuse scatterers are selected for channel computation. H_1 and H_2 are the height of Tx and Rx. Den_s and Den_d are the density of the diffuse and static scatters along the rail track. Parameters G_{LoS}, G_{DI} and G_{SD} are the channel gain of the line-of-sight (LoS) path, diffuse and static scatterers, and n_{LoS}, n_{DI} and n_{SD} denote the path loss exponents of the LoS path, diffuse and static scatterers.

2.2 Path Loss

Propagation path loss is an important parameter for the link estimation, required transmit power, and antenna gains. Path loss modeling have been studied by researchers for many years. The path loss for T2T urban viaduct channel is presented in Fig. 1. The path loss in fig. 2 shows a symmetric structure varying from the maximum value around 120 dB to the minimum value 70 dB, and then return to the maximum. This trend is corresponding to the movement of the Tx and Rx. They move from 3000 m away towards each other on the neighboring track. At the initial 3000 m, the path loss is the maximum 130 dB. As they approach each other, the path loss decreasingly reduces, and is down to 70 dB. At this time, they meet each other at a relative a very closed distance. Subsequently, they continue to move with their distance become large again, leading to the path loss to increase to 120 dB.

2.3 RMSDS

RMS-DS is used to describe the power dispersion over time (or delay). Large RMS-DS will result in the frequency-selective fading. The structure of RMSDS is also symmetric with regard to the time 15s. We will describe the first 15 s for short. In the fig. 3, the RMSDS firstly increases from 0.5 μs to the maximum value 1.7μs at 7.5 s. It then decreases to the minimum value when the two trains meet each other at 15 s. Not that the maximum RMSDS does occur at 0 s but at 7.5 s. We present the fitting distribution of the RMSDS in fig. 4. Weibull and Nakagami distribution are both fit for the RMSDS. The mean value of the RMSDS for the Weibull and Nakagami distribution are 1.37 μ s and 2.34 μ s. They both have a standard deviation 2.34 dB and 1.78 dB.

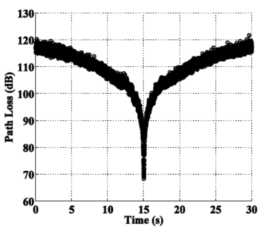


Fig. 2. Path loss for T2T urban viaduct channel

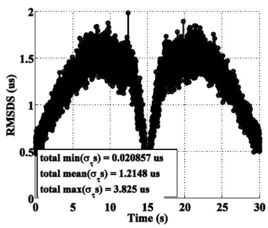


Fig. 3. RMSDS of T2T urban viaduct vs. time

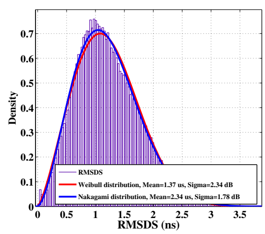


Fig. 4. Fitting distribution of the RMSDS

2.4 Ricean K Factor

The Ricean K-factor quantifies the small-scall fading characteristics of the radio propagation channel. Due to the statistically non-stationarity nature of the V2V channel, the Ricean K-factor will frequently vary with distance (or time). We use the method described in [14] to estimate the Ricean K-factor. Using this, k (in linear units) is expressed as follows:

$$\boldsymbol{K} = \frac{-2\mu_2^2+\mu_4-\mu_2\sqrt{\mu_2^2-\mu_4}}{\mu_2^2-\mu_4} \quad (1)$$

where μ_2 and μ_4 are the second and fourth moments of the received envelope. We estimate the Ricean *K*-factor using a 10 m sliding window with distance. The Ricean K factor of T2T urban viaduct channel are shown in Fig. 5.

Shown in fig. 5, the Ricean *K* factor varies vs. time ranging from -28 dB to 10 dB. It does not reflect an obvious trend like path loss, but fluctuating around 0 dB. In some areas, the Ricean *K* factor is extremely severe, down to -28 dB, which shows the weak condition of LoS path. We explore the fitting distribution for Ricean K factor. Weibull and Ricean distribution are the two suitable statistical distribution for Ricean K facor. The mean value of the two distribution are 2.19 us and 2.31 us with standard deviation 1.98 dB and 1.46 dB. Seen from Fig.6, the Ricean K factor follows Ricean distribution.

2.5 Envelope Covariance

The envelope covariance can be calculated from an important second-order statistic: the envelope autocorrelation functionρ(Δt), where Δt indicates time difference. For the T2T urban viaduct channel, ρ(Δt) determines the correlation of received envelope as a function of change in Rx motion time and is useful for studies in correlation properties, written as

$$\rho\,(\Delta t) = \frac{E\{[r(t)-E(r(t))][r(t+\Delta t)-E(r(t+\Delta t))]\}}{\sqrt{var[r(t)]var[r(t+\Delta t)]}} \quad (2)$$

where *r(t)* denotes the received envelope at time *t*, *E*[·] and *Var*[·] denote the expected value and the variance of [·] respectively. We define the correlation coefficient below 0.5 as non-correlation.

1) Received Envelope Large-Scale Autocorrelation

We retain the path loss variation in the received envelope to investigate the large-scale envelope autocorrelation in the T2T urban viaduct channel. In fig.7, the large-scale autocorrelation is shown. We plot a line of 0.5 to indicate the large-scale stationary time of T2T urban viaduct channel. The large-scale stationary time is 0.09 s. The constant velocity of train is 100 m/s.

2) Small-Scale Received Envelope Autocorrelation

We remove the path loss effect in the received envelope to find the small-scale envelope covariance in the received envelope. Shown in fig. 8, when the small-scale envelope covariance is 0.5, the stationary time is 0.008 s.

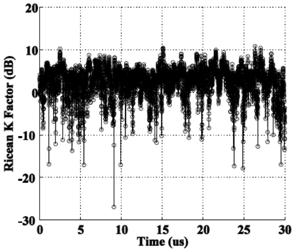


Fig. 5. Ricean *K* factor of T2T urban viaduct vs. time

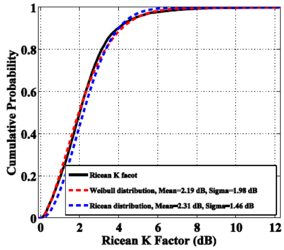


Fig. 6. Distribution of the Ricean *K* factor of T2T urban viaduct

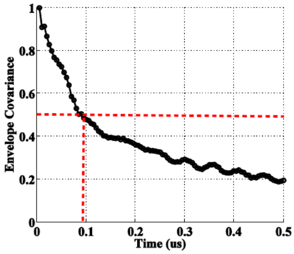


Fig. 7. Large-scale envelope covariance of T2T urban viaduct channel

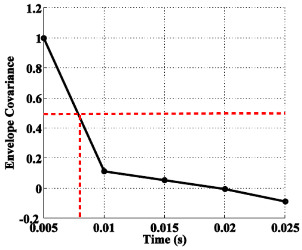


Fig. 8. Small-scale envelope covariance of T2T urban viaduct channel

2.6 TDL Model

We present the small-scale amplitudes updated vs. time in Fig. 9. The abscissa axis denotes multipath arriving time, the vertical axis is time variation, and the z axis is amplitude in dB scale. Seen from vertical axis, the tap amplitudes fluctuate rapidly with time. The later the path comes, the less frequently it is present. There is always a strong path around 1 μs, and the tap energy decreases exponentially along delay time.

The TDL model consists of the number of multipath components (MPCs), statistical distribution of MPCs and birth-and-death probability of MPCs. The tap number of the TDL model is computed by the maximum of the RMS-DS divided by the time resolution. The number of TDL model for T2T urban viaduct channel is ⌊3.825/1⌋+1=4. The energy of each tap is a relative value obtained from average PDPs. The relative tap energy of T2T urban viaduct channel is shown in Fig. 10. The energy of second tap is around 10 dB lower than that of first tap, and the third tap is 1 dB lower than that of the second tap. We can conclude that the tap energy does not decrease a lot, but 2 dB lower sequentially.

Table 2 shows the TDL model of T2T urban viaduct channel. The tap energy of from the 1^{st}-4^{th} tap is 0.796, 0.091, 0.065 and 0.048. Most of the energy are focused on the first tap. We use the Weibull distribution to characterize the small-scale tap variation. In the Weibull distribution, the shape factor b suggest the fading channel characteristics. In table 1, the first tap follows Ricean distribution to indicate a good LoS component, the 2^{nd}, 3^{rd}, and 4^{th} path are suffered a worth-than-Rayleigh channel with Weibull shape factor less than 2 [15], which shows the bad condition the MPCs.

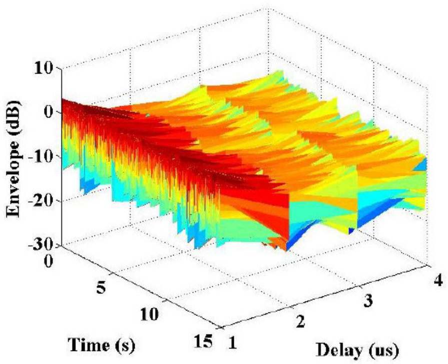


Fig. 9. Multipath amplitudes update vs. time

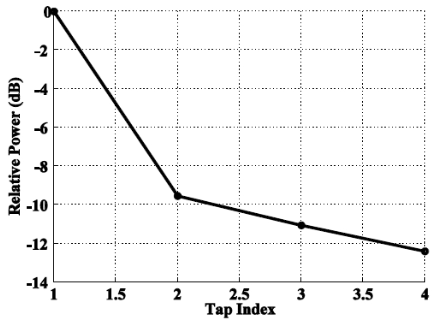


Fig. 10. Relative tap energy of T2T urban viaduct channel

Table 2. TDL Model of T2T Urban Viaduct Channel

Tap Index k	Energy	Shape Factor (b_k)
1	0.796	2.12
2	0.091	1.81
3	0.065	1.76
4	0.048	1.73

3 Conclusion

This paper presents the T2T GBSCM channel characterization in urban viaduct environment. In urban scenario, there are scatters (buildings and vegetation) providing reflected energy to the Rx. We characterize the path loss, RMS-DS, TDL model, Ricean K-factor, and large/small -scale envelope covariance. The path loss of T2T viaduct channel is symmetric with maximum path loss 120 dB, and the RMSDS follows Weibull and Nakagami distribution with mean value is 1.28μs. The Ricean K factor rapidly changes ranging from -28 dB to 10 dB. There are 4 taps in the TDL model. The first tap occupies the most energy.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant 61222105, the Fundamental Research Funds for the Central Universities under Grant 2010JBZ008, Beijing Municipal Natural Science Foundation under Grant 4112048, project of State Key Lab under Grant No. RCS2012ZT013, key Project for Railway Ministry of China under Grant 2012X008-A, interdiscipline cooperation projects of the New-Star of Science and Technology supported by Beijing Metropolis under grant no. xxhz201201, and the Key grant Project of Chinese Ministry of Education under Grant 313006.

References

[1] Accidents in Deciding Frequency by Cause, `http://safetydata.fra.dot.gov/officeofsafety/publicsite/Query/inctally1.aspx`

[2] Garcia, C.R., Lehner, A., Strang, T.: A broadcast vehicle-to-vehicle communication system in railway environments. In: 1st Annual International Symposium on Vehicular Computing Systems (2008)

[3] Strang, T., zu Hörste, M.M., Gu, X.G.: A railway collision avoidance system exploiting ad-hoc inter-vehicle communications and GALILEO, `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140`

[4] Garcia, C.R., Lehner, A., Stang, T., Röckl, M.: Comparison of collision avoidance systems and applicability to rail transport. In: ITST 2007: 7th International Conference on ITS, pp. 1–6 (2007)

[5] Garcia, C.R., Lehner, A., Strang, T., et al.: Channel model for train-to-train communication using the 400 MHz band. In: IEEE Vehicular Technology Conference, pp. 3082–3086 (2008)

[6] Lehner, A., Strang, T., García, C.R.: A Reliable Surveillance Strategy For an Automous Rail Collision Avoidance System, `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.4258`

[7] Lehner, A., Garcia, C.R., Wige, E., Stang, T.: A multi-broadcast communication system for high dynamic vehicular ad-hoc Networks. In: ICUMT 2009: International Conference, pp. 1–6 (2009)

[8] Lehner, A., Rico García, C., Strang, T., Heirich, O.: Measurement and Analysis of the direct train-to-train propagation channel in the 70cm UHF-band. In: Strang, T., Festag, A., Vinel, A., Mehmood, R., Rico Garcia, C., Röckl, M. (eds.) Nets4Trains/Nets4Cars 2011. LNCS, vol. 6596, pp. 45–57. Springer, Heidelberg (2011)

[9] Zhao, P.: The operation and reorganization of the high speed train. China Railway Publishing House (2009)

[10] Liu, P., Ai, B., Zhong, Z., Zhou, X.: A Novel Train-to-Train Communication Model Design Based on Multi-hop in High-Speed Railway. International Journal of Antennas and Propagation (2012) (accepted for publication)

[11] Liu, P., Ai, B., Zhong, Z., Zhou, X.: Outage Analysis of Train-to-Train Communication Model over Nakagami-m Channel in High-Speed Railway. International Journal of Antennas and Propagation (2012) (accepted for publication)

[12] Molisch, A.F.: A generic model for MIMO wireless propagation channels in macro- and microcells. IEEE Trans. Signal Processing 52(1), 61–71 (2004)

[13] Karedal, J., Tufvesson, F., Czink, N., Paier, A., Dumard, C., Zemen, T., Mecklenbrauker, C.F., Molisch, A.F.: A Geometry-Based Stochastic MIMO Model for Vehicle-to-Vehicle Communications. IEEE Trans. Wireless Comm. 8(7), 3646–3657 (2009)

[14] Gao, L.Y., Zhong, Z.D., Ai, B., Xiong, L.: Estimation of theRicean factor in K the high speed railway scenarios. In: Proc. CHINACOM 2010, Beijing, pp. 1–5 (August 2010)

A Cross-Layer Congestion Avoidance Routing Scheme in Vehicular Ad Hoc Networks

Chiao-Ling Yang, Yao-Hsing Chung, Chen-Yu Chen, and Chung-Ju Chang

Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan
original123466@gmail.com,
{eason.cm95g,duke953.cm96g,cjchang}@nctu.edu.tw

Abstract. We propose a cross-layer congestion avoid-ance routing (CCAR) scheme for VANETs based on IEEE 802.11p/WAVE (wireless access in vehicular environment) protocol. Due to highly dynamic network topology, the forwarders are selected one by one. The CCAR scheme first adjust the number of candidates adaptively so that senders have high probability to find a candidate as the forwarder. Then it dynamically adjusts the selection probability of each service channel (SCH) based on traffic load of SCHs to make channel load balanced. Finally, the CCAR scheme takes the number of co-channel senders, the loads of each possible forwarder, and the vehicular moving direction into account to determine the most appropriate route for senders. Simulation results show that the proposed CCAR scheme achieves lower packet delay (higher system uplink through-put) than the conventional GyTAR and AODV schemes by amounts of 95% and 95.7%, (8.4% and 19.2%,) respectively, when the number of vehicles is 150.

Keywords: vehicular ad hoc networks, wireless access in vehicular environment.

1 Introduction

Wireless access in vehicular environment (WAVE), which is a protocol designed for vehicular ad hoc networks (VANETs) on dedicated short-range communication(DSRC) band, is composed of the IEEE 802.11p [1] and the IEEE 1609 standard family. In IEEE 802.11p/WAVE, 75MHz spectrum at 5.9 GHz, which was allocated by United States Federal Communication Commission, is divided into seven channels: one control channel (CCH) and six service channels (SCHs). CCH is devoted to network control messages exchange, and SCHs are devoted to data packets exchange. Another important restriction in IEEE 802.11p/WAVE is that data packets exchange is only permitted within a WAVE-based basic service set (WBSS) [1]. A vehicle which established a WBSS cannot join a WBSS established by other vehicle simultaneously; in addition, a vehicle cannot join more than one WBSS. As a result, an efficient routing scheme in VANET is not only to consider highly dynamic network topology, but also to consider media access control (MAC) restrictions and problems in IEEE802.11p/WAVE standard. In this paper, we propose a cross-layer congestion avoidance routing (CCAR)

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_51, © Springer-Verlag Berlin Heidelberg 2014

scheme for VANETs based on IEEE 802.11p/WAVE. To avoid routing congestion, the proposed CCAR scheme controls the number of users on each SCH to reduce collision probability, and it selects the route path that has low collision possibility to ensure reliable transmission of data packets with low latency. The CCAR scheme contains three algorithms: 1) role determination (RD) algorithm, 2) service channel determination (SCHD) algorithm, and 3) forwarder selection (FS) algorithm. The RD algorithm is designed to prevent from no route problem by adaptively adjusting the number of senders and candidates so that senders have high probability of finding a candidate as the forwarder to deliver packets within their communication range. The SCHD algorithm aims to balance channel loads of each SCH. The SCHD algorithm dynamically adjusts the selection probability of each SCH based on the traffic loads of SCHs. The FS algorithm is proposed to avoid channel congestion which the sender may suffer from. The FS algorithm takes the number of co-channel senders, the loads of each possible forwarder, and the vehicular moving direction into account to determine the most appropriate route for senders. Simulation results show that the proposed scheme could reduce the average delay and improve the uplink throughput when the vehicular density is dense.

The rest of this paper is organized as follows. Section 2 describes the VANET for urban scenario and IEEE 802.11p/WAVE protocol. The proposed CCAR scheme is given in Section 3. Section 4 shows the simulation results and discussions. The conclusions and future works are given in Section 5.

2 System Models

We consider VANET in a two-way urban street, where each way has three lanes. The observation area is a rectangle region of the urban street, as illustrated in Fig. 1. Assume that there is an RSU located at each of vertex of the observation area, and two RSUs on the same road side are separated by L meters. Moreover, there are K vehicles in the observation area, each of which is equipped with a single-radio communication device. Each vehicle periodically generates data packets and uploads their data to the RSUs. Assume that data packets have delay requirement, denoted by $D*$. The communication ranges of both vehicles and RSUs are set to r. Every RSU periodically broadcasts messages to inform its existence to the vehicles that are within it communication range r.

We adopt IEEE 802.11p in the VANET we considered. It operates with IEEE 1609 standard family and develops a wireless access in vehicular environment (WAVE) protocol to support vehicular environments [1], [2].

DSRC Channels: The WAVE protocol defines seven channels in the dedicated short-range communication (DSRC) spectrum, where one of seven channels is the control channel (CCH) reserved for conveying network control messages, and other six channels are service channels (SCHs) used to exchange data packets. WAVE-mode short messages (WSMs), which are short, time-sensitive messages or system control messages, can be transmitted on both CCH and SCH. In the investigated VANET, WSMs are used to trigger forwarder negotiation processes and the number of available SCHs is denoted by N_S.

CCH and SCH Switching Model: Because each vehicle is equipped with a single-radio communication device, the vehicle needs to switch between the CCH and the SCHs. Therefore, the channel access time is divided into non-overlapping synchronization intervals with a fixed length of 100 milliseconds, comprising a CCH interval and a SCH interval. Both the CCH interval and the SCH interval are with a fixed length of 50 milliseconds. The multichannel architecture is specified in IEEE 1609.4 standard [2], and is shown in Fig. 2. The vehicles switch between CCH and SCH periodically and synchronously. During CCH interval, all vehicles must switch to CCH to monitor the CCH and exchange system control messages. During SCH interval, vehicles can switch to one of SCHs to exchange data packets with other vehicles/RSUs.

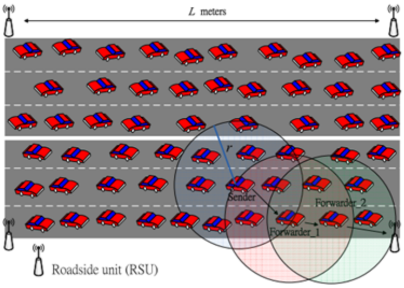


Fig. 1. Architecture of the investigated VANET

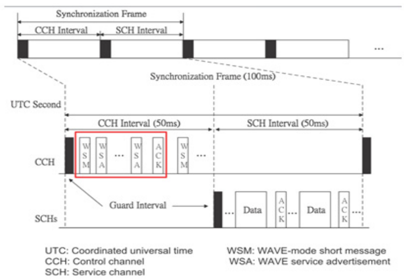


Fig. 2. WAVE 1609.4 multichannel structure

WBSS Operation Model: According to IEEE 802.11p/WAVE protocol [2], only the vehicles within the same WAVE basic service set (WBSS) are allowed to exchange data packets on a specified SCH without any association and authentication process. A vehicle that received the WSM then establishes a WBSS and periodically broadcasts the WAVE service advertisements (WSAs) on the CCH, where the WSAs advertise establishment of a WBSS and provide information. After collecting the WSAs, the vehicle who sent the WSM can join the interested WBSS by switching its operation band to the specified SCH of the WBSS on the subsequent SCH interval. Note that a vehicle cannot simultaneously establish a WBSS and join a WBSS established by other vehicle, and a vehicle cannot join more than one WBSS neither. Three types of node are defined in the proposed CCAR scheme:

•Sender: A node which has data to send broadcasts WSM.
•Candidates: The nodes which may play role of forwarding data packets for the sender.

There are two types of candidate depending on whether it has received WSM or not:

– Infeasible candidate: The candidate that does not receive the WSM is a Infeasible candidate.

– Feasible candidate: The candidate that received the WSM and replied with the WSA is a feasible candidate for the forwarder of the sender, "feasible candidate" for short.

•Forwarder: The node to which the sender will transmit data packets on the subsequent SCH interval. The feasible candidate that received the acknowledgement (ACK) from a sender is called "forwarder of sender".

3 Cross-Layer Congestion Avoidance Routing (CCAR) Scheme

In this section, we propose a CCAR scheme. The CCAR scheme mainly contains three algorithms: a role determination (RD) algorithm, a service channel determination (SCHD) algorithm, and a forwarder selection (FS) algorithm.

■ Operation Procedure of CCAR Scheme

At the beginning of CCAR scheme, the vehicles perform the RD algorithm and will be cataloged into either of two types of nodes: candidates and sender. After determining the role, candidates and the sender perform respective operation procedures. During the CCH interval, both candidates and the sender switch to the CCH to exchange information. The sender and the candidates monitor the CCHin the CCH interval and count the WSAs and corresponding ACKs which are sent by other feasible candidates and senders, respectively, as shows in Fig. 2. When the SCH interval begins, the sender and the chosen forwarder switch to the specified SCH to exchange data.

At the beginning of the CCH interval, the candidate performs the SCHD algorithm to select an appropriate SCH that has low collision probability as its specified SCH. After the SCH is determined, the candidate monitors the CCH and waits for a WSM. When the candidate receives a WSM, it becomes a feasible candidate. The feasible candidate initiates a WBSS and replies a WSA to the sender. Then, the candidate expects an ACK that acknowledges that the sender joins one of the WBSSs. If the feasible candidate receives the expected ACK, it stops periodically broadcasting the WSA and checks the ACK. If the ACK is the reply to its WSA, it means that the sender joins the WBSS of the feasible candidate, and the feasible candidate becomes the forwarder of the sender. The feasible candidate will continues monitoring the CCH. When the SCH interval begins, the candidate switches to the selected SCH. During the SCH interval, the forwarder receives data which are sent from the senders.

The sender broadcasts WSM to inform its one-hop nodes that it has data to send. After receiving the WSAs from the feasible candidates, the sender performs the FS algorithm to select the most appropriate forwarder. After determining the forwarder, the sender will continue monitoring the CCH. When the SCH interval begins, the sender switches to the specified SCH and transmits packets to the forwarder on the specified SCH within the SCH interval.

■ Role Determination Algorithm

When the number of vehicles in the observation area becomes heavy, the situation that every vehicle has packets to send would happen, which means that every vehicle will become sender and no one plays the candidate/forwarder to deliver packets for others. In this situation, the vehicles that are not within the communication range of RSUs could hardly find a route path to RSUs, especially for ones which are far from RSUs. We call this situation no route problem. In order to solve no route problem, we design a candidate threshold in the RD algorithm to adaptively adjust the number of senders and candidates. Assume that the candidate threshold of vehicle i is denoted by T^*_i, where $0 \le T^*_i \le 1$. After T^*_i is determined, vehicle i randomly chooses a number t from [0, 1]; vehicle i will play the role candidate if $t \le T^*_i$, and otherwise vehicle i will play the role sender. Because vehicles which are far from RSUs have to take longer time to deliver their packets to RSUs due to high number of hops in route path, these vehicles are assigned a lower candidate threshold. In order to avoid that the vehicles which are closer to RSUs would usually play candidates and lose the chance of sending their own packets, the design of the candidate threshold also takes the number of times that a vehicle continuously plays a candidate into account for fairness consideration. If a vehicle plays candidate in a row, it would be assign a low candidate threshold. Note that if a vehicle does not have data packets to send, its candidate threshold is set to 1. Based on designed principles described above, candidate threshold T^*_i of vehicle i is defined as

$$T_i^* = \begin{cases} (\frac{1}{1+h_{v,i}})^{(1+t_{c,i})}, & \text{if } N_{v,i} > 0, \\ 1, & \text{if } N_{v,i} = 0, \end{cases} \tag{1}$$

where $N_{v,i}$ is the number of packets that vehicle i possesses, $h_{v,i}$ denotes the number of hops between vehicle i and its nearest RSU, and $t_{c,i}$ denotes the number of times that vehicle i continuously plays a candidate. The initial values of $t_{c,i}$ is set to zero, $1 \le i \le K$. Vehicle i in- creases $t_{c,i}$ by one when it determines to play a candidate, i.e. $t_{c,i} = t_{c,i} + 1$, and vehicle i resets $t_{c,i}$ to zero when it determines to play a sender, i.e. $t_{c,i} = 0$. By updating $t_{c,i}$, vehicles will take turns to be senders and candidates.

■ Service Channel Determination Algorithm

At the beginning of CCH interval, each candidate performs the SCHD algorithm to determine an appropriate SCH for itself. In the SCHD algorithm, candidates choose their specified SCHs simultaneously. That each candidate selects the SCH with lowest load as its specified SCH may lead to candidates gather at one of SCHs. It makes the SCH overload, and causes high collision probability. Therefore, we propose a way that candidates choose their specified SCH with probability. We denote jth feasible candidate of the considered sender as "feasible candidate j". Firstly, feasible candidate j checks whether its SCH has been determined in the previous frame. If it had not determined SCH in the previous frame, it determines its SCH according to SCH selection probability. If feasible candidate j had determined SCH in the previous frame, it has to calculate the successful reception ratio of previous frame, denoted by R_j, and compare R_j with the required successful reception ratio, denoted by $R*$. If the value of R_j is lower than R^*, feasible candidate j determines new SCH based on SCH selection probability. If the value of Rj is equal to or greater than R^*, feasible candidate j uses the same SCH as previous frame.

SCH Selection Probability: Let $p_{j,m}$ be the probability that feasible candidate j will select SCH m as its SCH, which is defined as

$$p_{j,m} = \frac{\frac{1}{1+v_{j,m}}}{\sum_{k=1}^{N_S} \frac{1}{1+v_{j,k}}}, 1 \leq j \leq |\psi|, \tag{2}$$

where $v_{j,m}$ denotes the number of SCH m within the communication range of feasible candidate j in the previous frame and ψ denotes the set of feasible candidates of the considered sender. After deriving the $p_{j,m}$, m = 1, . . . , N_S , feasible candidate j chooses a random number, denoted by H, which is uniformly distributed in [0,1]. According to the value of η, feasible candidate j can determine its SCH, denoted by sj.

Successful Reception Ratio: In the CCAR scheme, we assume that ACKs sent from senders contain information about the number of packets sender has to send. The number of packets excepted to be receive by feasible candidate j in the previous frame, denoted by r_j^E , can be known by feasible candidate j. The R_j can be expressed as

$$R_j = \frac{r_j^S}{\max(r_j^E, 1)}, \tag{3}$$

where r_j^S denotes the number of packets successfully received by feasible candidate j in the previous frame.

■ Forward Selection Algorithm

During the CCH interval, senders perform the FS algorithm to select an appropriate forwarder for the sender to ensure reliable transmission of packets with low latency. The cross-layer design concept of FS algorithm is that the sender uses the information provided by MAC layer to determine route path. In the FS algorithm, we also consider the balance of load between every feasible candidate of the considered sender and the effect of moving direction of each feasible candidate. Design concept of the FS algorithm is to consider collision avoidance, load balance, and delay reduction simultaneously. According to these three considerations, we design a cost function, which will be described in section 3-D.1. According to the information contained in WSA, the considered sender can calculate the cost of choosing this feasible candidate as its forwarder. When the WSA receiving timer is timeout, the considered sender calculates the cost of each feasible candidate, and then selects the feasible candidate with minimum cost as its forwarder. If the considered sender did not receive any WSA after the WSA receiving timer is timeout, the considered sender should broadcast WSM again.

Cost Function: According to the WSAs, the considered sender can know the number of hops between feasible candidate j and the nearest RSU of feasible candidate j, denoted by h_j. Let h 0 denote the number of hops between the considered sender and its nearest RSU, which is defined as

$$h_0 = \min(h_j) + 1. \tag{4}$$

In order to avoid that packets will be forwarded to inappropriate forwarder, which causes redundant packet transmissions, we let the cost is infinity if hj is equal to or larger than h 0 . The cost of feasible candidate j, denoted by Cj , is defined as

$$C_j = \begin{cases} w_1 \times V_j + w_2 \times I_j + w_3 \times H_j, & \text{if } h_j < h_0, \\ \infty, & \text{if } h_j \geq h_0, \end{cases} \tag{5}$$

where w 1 , w 2 , and w 3 are weight factors, Vj denotes a collision factor, Ij denotes a unfairness factor, and Hj denotes a delay factor. The contents of collision factor, unfairness factor, and delay factor will be described below.

(A) Collision Factor: Vj represents the possibility of collision of feasible candidate j. According to the WSAs, the considered sender can know sj , where j = 1, . . . , | ψ | . Before sending an ACK, the considered sender has heard the ACKs sent from other senders within its communication range. From the ACKs, the considered sender can know which SCH the other senders within its communication range select. The number of senders already select SCH m within the communication range of the considered sender is denoted by a m. As the number of contestant increases, the possibility of collision increases rapidly. If the possibility of collision is high, the cost will also be high. We defined the collision factor as

$$V_j = \frac{\exp^{a_{s_j}}}{\sum\limits_{k=1}^{N_S} \exp^{a_k}}. \tag{6}$$

(B) Unfairness Factor: Ij represents the unfairness index of feasible candidate j. Here, "Fairness" is defined as distributing equal load to all feasible candidates. According to the WSAs, the considered sender can know the number of packets that each feasible candidate will have. Let Nj be the number of packets that feasible candidate j may have, which is the sum of the number of packets feasible candidate j has to receive and the number of packets feasible candidate j already had. Let Fj be the fairness index of feasible candidate j, which is used to estimate the degree of fairness. We set the fairness index, by [3], as

$$F_j = \frac{[(\sum\limits_{k=1}^{|\psi|} N_k) + N_0]^2}{|\psi| \times [\sum\limits_{\substack{k=1 \\ k \neq j}}^{|\psi|} (N_k)^2 + (N_j + N_0)^2]}, \tag{7}$$

where N0 denotes the number of packets the considered sender has to send. The value of Fj is between 0 to 1. If the value of fairness index is close to 1, it means that the load between feasible candidates in ψ are very close. In order to balance the load of feasible candidates, the considered sender prefers to select the feasible candidate with lower load as its forwarder. Therefore, the larger the fairness index of feasible candidate j is, the less the cost of feasible candidate j is. An unfairness factor is defined as

$$I_j = 1 - F_j. \tag{8}$$

(C) Delay Factor: Hj represents the delay index of feasible candidate j. According to the WSAs, the considered sender can know the number of hops between each feasible candidate and its nearest RSU. The fewer the number of hops is, the shorter the end-to-end delay is. Furthermore, the considered sender also can know the index which represents whether feasible candidate j is approaching its nearest RSU, denoted by λj from the WSA sent from feasible candidate j. When feasible candidate j is leaving its nearest RSU, λj is set to 1. When feasible candidate j is approaching its nearest RSU, λj is set to -1. Therefore, the delay factor is defined as

$$H_j = \begin{cases} \lambda_j \times \frac{1}{h_j}, & \text{if } \lambda_j = -1, \\ \lambda_j \times (1 - \frac{1}{h_j}), & \text{if } \lambda_j = 1. \end{cases} \tag{9}$$

4 Simulation Results and Discussions

In the simulations, parameters of the considered VANETs are set to be compatible with IEEE 802.11p and IEEE 1609 standard family [1], [2]. Each vehicle is assigned a velocity with a uniform distribution from 40 to 60 km/hr at the initiating of the simulation. Each vehicle is allowed to change lane and change velocity according to the distance between itself and the vehicle in front during simulation. All vehicles generate data packets with a rate of 5 packets/sec and a data packet size of 512 bytes. The destination of each vehicle is one of the RSUs. In addition, the RTS/CTS method is turned off in the simulation, because the data packet size is small, and the RTS/CTS method makes the routing overhead large. In order to let each node have chance to transmit their data packets, each node can only transmit one data packet for successful contention. Due to there are two dedicated SCHs, the number of available SCHs, N_S, is set to four. The weighting factors in the cost function, $(\omega_1; \omega_2; \omega_3;)$, are set to (0.4; 0.4; 0.2). Analysis for these weighting factors will be the future work to determine the balance between channel collision, node load, and moving direction for delivery efficiency. Each simulation lasts for 300 seconds, and the simulation results are the average of five runs.

The proposed CCAR scheme is compared with the adhoc on-demand distance vector (AODV) routing scheme [5] and the improved greedy traffic aware routing (GyTAR) scheme [6]. The AODV routing scheme is a conventional topology based and reactive routing scheme. It creates a route between two nodes only when the route is requested by the source node. The GyTAR scheme is a geographical routing scheme for VANETs in city scenarios. It dynamically selects the junction, and then uses the improved greedy strategy to select forwarder to forward data packets between two junctions. In this simulations, the connections between any two vehicles are based on the constrain, only the vehicles within the same WBSS are allowed to exchange data packets on a specified SCH. Fig. 3 shows the mean number of successful connections per frame of the three schemes. It can be seen that the CCAR scheme has the highest mean number of successful connections, and it improves the mean number of successful connections by amounts of 180% and 336% compared to the GyTAR scheme and the AODV scheme, respectively, when the number of vehicles is 150. It is because the CCAR scheme employs the RD algorithm to control the number of senders and candidates. The RD algorithm of the CCAR scheme makes senders have high probability of finding a route path to deliver their packet. In contrast, the AODV scheme and the GyTAR scheme do not have the capability to adaptively adjust the number of senders. In the GyTAR and AODV schemes, if a vehicle has data packets, it definitely acts as a sender. As a result, there will be few forwarders because most of vehicles become senders rather than forwarder. Then the problem of that a sender does not receive the WSA sent from its chosen forwarder occurs frequently. This problem results in unsuccessful connection.

Fig. 4 shows the average packet delay of the three schemes, where the average delay is defined as the average of the time between the time when a packet generating from the vehicle and the time when the packet received by an RSU. The average delay of the CCAR scheme is lower than that of the GyTAR and AODV schemes by an amount of about 95.7 % when the number of vehicles is 150. It is because the senders in the CCAR scheme perform the FS algorithm to select their forwarders, the FS algorithm considers the possibilities of collision of each SCH and the loads and moving directions of each feasible candidate simultaneously. Therefore, the senders choose the feasible candidate with lower collision possibility, less data packets in buffer, and the same moving direction as packet forwarding direction as their forwarders. It can reduce the packets retransmission times, the time data packets stay in buffer, and the redundant forwarding times.

Fig. 5 shows the system uplink throughput of the three schemes. The CCAR scheme has the largest uplink throughput among all schemes, and the CCAR scheme improves the uplink throughput by amounts of 19.2% and8.4% compared to the AODV scheme and the GyTAR scheme, respectively, when the number of vehicles is 150. In the CCAR scheme, the candidates perform SCHD algorithm to determine their SCHs. The candidates select their SCHs according to the selection probability of each SCH, which can avoid the ping-pong effect and balance the loads of SCHs. Moreover, the senders in the CCAR scheme determine their forwarders, which have low number of hops, low collision possibility, and less data packets in buffer. It results in lower packet delay. Hence, within the same time, the CCAR scheme can transmit more packets to RSUs than the GyTAR scheme or AODV scheme.

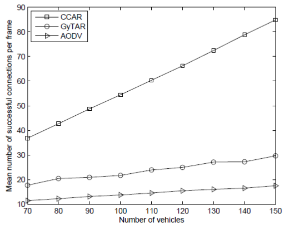


Fig. 3. The mean number of successful connections

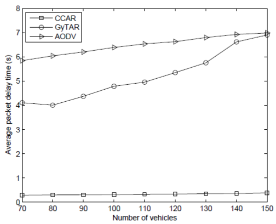


Fig. 4. The average packet delay

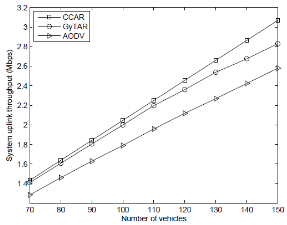


Fig. 5. The system uplink throughput

5 Conclusions and Future Works

A cross-layer congestion avoidance routing (CCAR) scheme is proposed for uplink vehicular ad hoc networks (VANETs), and based on wireless access in vehicular environment (WAVE) protocol. The CCAR scheme determines the appropriate route path with low packet dropping ratio and low end-to-end delay by considering the collision possibility in the MAC layer, the load balances of both SCHs and forwarders, and the moving directions of vehicles. We have shown that the average delay of the CCAR scheme is lower than that of the GyTAR and AODV schemes by an amount of about 95.7 % when the number of vehicles is 150. Moreover, the CCAR scheme improves the uplink throughput by amount of 19.2% and 8.4% compared to the AODV scheme and the GyTAR scheme, respectively, when the number of vehicles is 150. In the future, we will extend the research from fixed transmission range to dynamical transmission range.

References

[1] IEEE Standard for Information Technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifi cations Amendment 6: Wireless Access in Vehicular Environments, IEEE Std 802.11p-2010 (July 2010)

[2] IEEE Standard for Wireless Access in Vehicular Environments (WAVE)–Multi-channel Operation, IEEE Std 1609.4-2010 (February 2011)

[3] Jain, R., Chiu, D., Hawe, W.: A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. DEC Research Report TR-301 (September 1984)

[4] Jiang, D., Chen, Q., Delgrossi, L.: Optimal data rate selection for vehicle safety communications. In: Proceedings of 5th ACM International Workshop on Vehicular Internetworking, pp. 30–38 (2008)

[5] Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: Proceedings of 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 90–100 (February 1999)

[6] Jerbi, M., Senouci, S.-M., Meraihi, R., Ghamri-Doudane, Y.: An improved vehicular ad Hoc routing protocol for city environments. In: IEEE International Conference on Communications, pp. 3972–3979 (June 2007)

Theoretical Aspects of Petri Nets Decomposition Based on Invariants and Hypergraphs

Remigiusz Wiśniewski, Łukasz Stefanowicz,
Arkadiusz Bukowiec, and Jakub Lipiński

University of Zielona Góra, Institute of Computer Engineering and Electronics,
ul. Licealna 9, 65-417 Zielona Góra, Poland
{r.wisniewski,a.bukowiec}@iie.uz.zgora.pl,
{l.stefanowicz,j.lipinski}@weit.uz.zgora.pl

Abstract. Two methods of Petri nets decomposition into State Machine Components (SMCs) are shown in the paper. The first one bases on the well-known algorithm of place invariants (p-invariants) calculation. The second method applies hypergraph theory and computation of exact transversals. The aim of the paper is theoretical analysis of the effectiveness of presented methods. We show, that despite the high popularity, the achieved results generated by p-invariants are not always correct and spurious components ought to be eliminated. Furthermore, the effectiveness of the application of hypergraphs into Petri net decomposition is analysed.

Keywords: Petri net, decomposition, State Machine Component (SMC), linear programming, p-invariant, hypergraph, exact transversal.

1 Introduction

Application of graphical representation of algorithm is very conformable [1-5] method of specification of dedicated binary controllers. In this case Petri nets (PNs) [6-8] are one of the most adequate methods for formal design of such controllers [9]. It gives easy way for representation of concurrent processes and additionally there could be applied mathematical algorithms for formal analysis and verification of the designed model [10-12]. There are also several algorithms of direct [13, 14] or distributed [15, 14] synthesis of Petri net model into programmable devices.

In the paper we will focus on the comparison of the theoretical aspects of Petri net decomposition. Two different decomposition methods are shown in the paper. Algorithms are theoretically analyzed due to computational complexity, effectiveness and validity of results.

2 Main Definitions and Preliminary Notation

Definition 1. *A Petri net* [6, 7] *is a 4-tuple $PN=(P, T, F, M_o)$ where P is a finite set of places, T is a finite set of transitions, $F \subseteq (P \times T) \cup (T \times P)$ is a finite set of arcs, M_o is an initial marking.*

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_52, © Springer-Verlag Berlin Heidelberg 2014

State of a net is called *marking*. Marking can be also seen as a distribution of the tokens in the net places. If a place contains one or more tokens, it is called as *marked place*. Movement between markings are done after *firing* of the transition. Petri net is *live* if from any marking of a Petri net it is possible to fire any transition in a net by sequence of firings of other transitions. A place of a net is *safe* if there is no reachable state that contains more that one token in this place. A Petri net is safe if each place in the net is safe. The net is *conservative* if all the markings contain the same number of tokens. Petri net can be represented by the *incidence matrix C*. Rows of the matrix correspond to places, while columns refer to transitions of a Petri net. Cell c_{ij} of matrix C is connected with place p_i and transition t_j in such a way, that $c_{ij} = c_{ij}^{+} - c_{ij}^{-}$ where (for safe Petri nets):

$$c_{ij}^{+} = \begin{cases} 1, (t_j, p_i) \in F \\ 0, otherwise \end{cases}, \quad c_{ij}^{-} = \begin{cases} 1, (p_i, t_j) \in F \\ 0, otherwise \end{cases}.$$

An exemplary Petri net is shown in Fig.1 (a). The net contains six places and three transitions. It is live, conservative and safe. The set of reachable markings contains three states reached by cyclic firing of successive transitions, as it is shown in the Fig.1 (b): $M_0 = (a,b,f) \xrightarrow{t_1} M_1 = (b,c,d) \xrightarrow{t_2} M_2 = (a,c,e) \xrightarrow{t_3} M_0$.

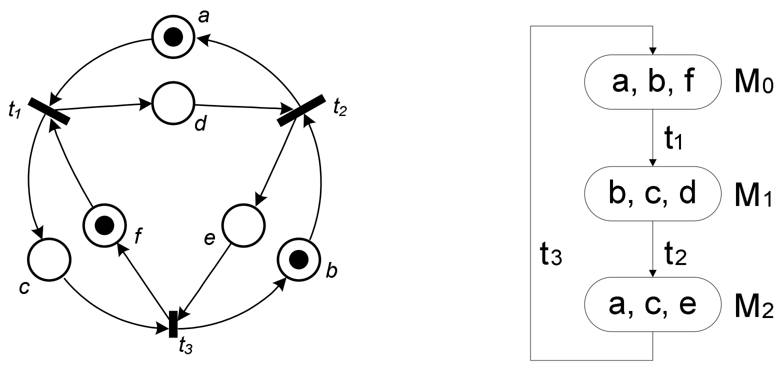


Fig. 1. Petri net PN_1 (a) and its reachable markings (b)

Definition 2. *A state machine component (SM-Component, SMC) of a Petri net PN is defined* [6] *as a strongly connected subnet PN' generated by places in PN, such that all input and output transitions of places in PN' and their connecting arcs belong to PN' and each transition of a subnet PN' has exactly one input and one output arc.*

Each SMC forms an independent sequential system. It is achieved as a result of a *decomposition* of a net. Such a process allows splitting a controller described by a Petri net into concurrent automata [14, 16, 17] in order to FPGA implementation [2, 8], effective states encoding [18] or advanced analysis of a net [7].

Petri net PN_1 can be decomposed into three State Machine Components: $S_1=\{a,d\}$, $S_2=\{b,e\}$ and $S_3=\{c,f\}$. Let us point out, that it is the only one way of the decomposition of PN_1.

Definition 3. *Place invariant is a vector* $y \in \mathrm{IN}_0^N$ *where* $y \neq 0$ *and* $yC = 0$.

Definition 4. *Hypergraph H [19] is defined by a pair H=(V,E), where* $V=(v_1,\ldots,v_n)$ *is an arbitrary, non-empty set of vertices and* $E=\{E_1,\ldots,E_m\}$ *is a set of hyperedges, subsets of V.*

Definition 5. *An exact transversal [20] of a hypergraph H = (V,E) is a set* $D \subseteq V$ *that has exactly one intersection with every edge of a hypergraph:* $\forall E_i \in E : |E_i \cap D| = 1$

3 Decomposition of Petri Nets

In this section we briefly describe two different ways of Petri net decomposition. Each method is theoretically analysed by an example. Let us point out that we present pure methods, without any improvements (such as approximation or computation of the reduced set of solutions). Our goal is to indicate strong points and show weakness of classical invariant and hypergraph algorithms that are applied into decomposition of Petri nets.

3.1 Place Invariant Computation

Place invariants (p-invariants) are widely used for Petri nets decomposition and correspond to the conservation of tokens in subsets of places. There are several methods of p-invariants computation [6, 7, 21-23]. In the paper we chose the classical and well-known Matinez-Silva algorithm [10] as a representative.

Consider Petri net shown in the Fig. 1 (a). Let us now decompose PNi by obtaining p-invariants. At the beginning, the initial unit matrix $Q=[D|C]$ is formed (initially, D is equal to an identity matrix I_n). Next, the algorithm repeats the same procedure for each column (transition) t_j [10, 21]:

1. Find row pairs that are annul (their sum is equal to 0) the j-th column of C and append it to the matrix Q.
2. Delete all rows of Q in which the intersection with j-th column is not 0.

Obviously, the main advantage of the presented method is computational complexity of the outer loop, which is bounded by a polynomial in the size of the number of transitions. However, the number of rows that are pairwise annul can be exponent. Therefore, the whole algorithm is bounded by an exponential in the size of the number of transitions and places of the initial net. Furthermore, we show that obtained solution is not always correct.

Figure 2 presents the result of application of Martinez-Silva algorithm to Petri net PN_1. Final matrix contains five p-invariants, represented by rows of a matrix D. Clearly, the first p-invariant $y_1=(1,1,1,0,0,0)$ does not refer to a proper SMC.

$$Q = \begin{array}{c} \begin{array}{ccccccccc} a & b & c & d & e & f & t_1 & t_2 & t_3 \end{array} \\ \left[\begin{array}{cccccc|ccc} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \text{invariants} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{array}$$

Fig. 2. The output matrix of Martinez-Silva algorithm for Petri net PN_1

It contains two initially marked places: *a* and *b* which cannot belong to the same SMC. Thus, y_1 represents a spurious state in the net [16, 10].

After the detection and reduction of the spurious invariant, the net can be successfully decomposed into three SMCs: $S_1=\{a, d\}$ (refers to the invariant y_3), $S_2=\{c, f\}$ (y_4) and $S_3=\{b, e\}$ (y_5). Spurious invariants are relatively easy to detect, because they contain more than one initially marked place. Unfortunately, they have an obvious impact on the computational complexity of an algorithm.

3.2 Hypergraph Exact Transversal Computation

Application of hypergraph exact transversals into Petri net decomposition is relatively new idea, initially proposed in [14]. The method can be basically split into two main parts. At the beginning the concurrency hypergraph for the net is formed. Obviously, such a hypergraph refer to a reachability set of a Petri net [14]. Next, exact transversals in the concurrency hypergraph are calculated. Obtained sequentiality hypergraph contains all possible State Machine Components of the initial Petri net. The main advantage of this method is its reliability. In opposite to invariants computation, achieved result is always correct. However, the computational complexity of the whole algorithm can be exponential and in some cases determination of SMCs takes a long time or is even not possible.

Let us again decompose Petri net PN_1, but this time we apply hypergraph theory. At the beginning the concurrency hypergraph is formed. Such a structure can be simply obtained by analysis of concurrency relation between particular places in the net. Description of the concurrency hypergraph formation can be found in [17]. Figure 3 presents an incidence matrix of a concurrency hypergraph for the net PN_1. Each edge represents concurrency relation between places.

The complexity time of subsequent hyperedges computation (which refer to the concurrency relation) is bounded by a polynomial in the size of the number of places [17]. Unfortunately, the total number of hyperedges can be exponential. Therefore, similarly as in the case of invariants, the computational complexity of the whole algorithm turns to exponential.

$$A_{Hc} = \begin{array}{c} \begin{matrix} a & b & c & d & e & f \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Fig. 3. An incidence matrix of a concurrency hypergraph for net PN_1

At the next step, State Machine Components are calculated. Each SMC refer to an exact transversal in the concurrency hypergraph. The computation of subsequent SMC is bounded by a polynomial [14] in the size of the number of hypergraph vertices (which refer to net places) and edges (concurrency relations). Again, the total number of obtained SMCs can be exponential.

There are four exact transversals in the concurrency hypergraph of the net PN_1: $D_1=\{a, d\}$, $D_2=\{b, e\}$, $D_3=\{c, f\}$ and $D_4=\{d, e, f\}$. It means, that achieved result is the same as in the case of p-invariants computation. However, exact transversals do not generate spurious states and all indicated SMCs are correct.

4 Conclusions

Two different Petri net decomposition methods were presented in the paper. Each algorithm has strong and weak points. Both methods are bounded by an exponential in the size of the number of inputs (places and transitions) of the initial net. Fortunately, most of live and safe Petri nets, that describe real devices can be successfully decomposed within reasonable time [14, 16].

Application of the linear programming into Petri net decomposition implies additional spurious states which have to be detected and eliminated. Moreover, spurious invariants extends the size of the unit matrix, which may influence into the memory usage of an algorithm. On the other side, the method permits to generate invariants directly from the structure of the net, without any additional steps.

Decomposition of Petri nets through computation of exact transversals of hypergraphs always lead to the proper results [14]. Subsequent SMCs are obtained in a polynomial time, however the total number of components can be exponent. Moreover, the method requires computation of the concurrency hypergraph, which is also bounded by an exponential in the size of the number of inputs of the net.

Concluding, we would like to point out that both methods have their advantages and disadvantages. Computation of p-invariants is relatively fast, but may be insufficient, especially in case of nets containing a lot of spurious states. Application of hypergraph exact transversals calculation always result in proper results. However, the method bases on the two-step decomposition and the concurrency hypergraph has to be obtained initially.

References

1. Gomes, L., Costa, A.: From use cases to system implementation: Statechart based co-design. In: Proc. 1st ACM & IEEE Conf. Formal Methods and Prog. Models for Codesign, Mt. Saint-Michel, France, pp. 24–33. IEEE Comp. Soc. Press (2003)
2. Milik, A., Hrynkiewicz, E.: Synthesis and implementation of reconfigurable PLC on FPGA platform. Int. Journal of Electronics and Telecom. 58(1), 85–94 (2012)
3. Łabiak, G.: From UML statecharts to FPGA - the HiCoS approach. In: Proc. Forum on Spec. & Design Languages, Frankfurt, Germany, ECSI, pp. 354–363 (2003)
4. Doligalski, M.: Behavioral specification diversification for logic controllers implemented in FPGA devices. In: Proc. of FPGAworld 2012, pp. 6:1-6:5. ACM, New York (2012)
5. Grobelna, I.: Formal verification of embedded logic controller specification with computer deduction in temporal logic. Przegląd Elektrotechniczny 87(12a), 40–43 (2011)
6. Murata, T.: Petri nets: Properties, analysis and applications. Proc. IEEE 77, 541–580 (1989)
7. Karatkevich, A.: Dynamic analysis of PNet-based discrete systems. Springer (2007)
8. Gomes, L., Paulo, J., Costa, A.: Structuring Mechanisms in Petri Net Models: From specification to FPGA based implementations. In: Design of Embedded Control Systems, pp. 153–166. Springer (2005)
9. Marranghello, N., Mirkowski, J., Bilinski, K.: Synthesis of synchronous digital systems specified by Petri nets. In: Hardware Design and Petri Nets, pp. 129–150. Kluwer (2000)
10. Martinez, J., Silva, M.: A simple and fast algorithm to obtain all invariants of a generalized Petri net. Selected Papers from the European Workshop on App. and Theory of Petri Nets, pp. 301–310. Springer, London (1982)
11. Bazydlo, G., Adamski, M.: Specification of UML 2.4 HSM and its computer based implementation by means of Verilog. Przegląd Elektrotechniczny 87(11), 145–149 (2011)
12. Karatkevich, A., Gratkowski, T.: Analysis of the operational Petri nets by a distributed system. In: Proc. of TCSET, Lviv, Ukraine, pp. 319–322 (2004)
13. Tkacz, J., Adamski, M.: Macrostate encoding of reconfigurable digital controllers from topological petri net structure. Przegląd Elektrotechniczny 88(8), 137–140 (2012)
14. Wiśniewska, M.: Application of Hypergraphs in Decomposition of Discrete Systems. LNCCS, vol. 23. University of Zielona Gora Press, Zielona Gora (2012)
15. Bukowiec, A., Mróz, P.: An FPGA synthesis of the distributed control systems designed with Petri nets. In: Proc. of IEEE 3rd Int. Conf. on Networked Embedded Systems for Every Application NESEA 2012, Liverpool, UK (2012)
16. Dylewski, R., Adamski, M., Jablonski, J.: Application of linear programming for analysis of petri net. In: Proc. of KNWS 2011, Karpacz, Poland, pp. 267–273 (2011)
17. Wiśniewski, R., Wiśniewska, M., Adamski, M.: A polynomial algorithm to compute the concurrency hypergraph in Petri nets. MAaM 58(7), 650–652 (2012)
18. Carmona, J., Cortadella, J.: State encoding of large asynchronous controllers. In: DAC, pp. 939–944 (2006)
19. Berge, C.: Hypergraphs: Combinatorics of Finite Sets. North-Holland (1989)
20. Eiter, T.: Exact transversal hypergraphs and application to Boolean functions. J. Symbolic Computation 17(3), 215–225 (1994)
21. Lin, C., Zhang, T.: An Algorithm for Computing S-invariants for High Level Petri Nets. Fachberichte Informatik. Univ. (1991)
22. Zaitsev, D.: Decomposition-based Calculation of Petri Net Invariants. In: Proc. of the 25th Int. Conf. on Application and Theory of Petri Nets, June 21-25, pp. 79–83 (2004)
23. Esparza, J., Silva, M.: Top-down synthesis of live and bounded free choice nets. In: Rozenberg, G. (ed.) APN 1991. LNCS, vol. 524, pp. 118–139. Springer, Heidelberg (1991)

A Case Study of Embedded Software Development Based on Aspect-Oriented Programming

Mingrui Zhang[1], Wansheng Miao[2], Xin Zhang[2], and Ruonan Rao[1]

[1] School of Software, Shanghai Jiao Tong University, Shanghai, P.R. China
mingrui.sjtu@gmail.com, rnrao@sjtu.edu.cn
[2] China Aeronautical Radio Electronics Research Institute, Shanghai, P.R. China
{miao_wansheng,zhang_xin}@careri.com

Abstract. The aspect-oriented programming (AOP) is gaining popularity for it offers a means to encapsulate crosscutting concerns in the development of complex software systems. The power of AOP enables more degree of separation-of-concerns which leads to more understandable, maintainable and extensible software. However, challenges arise when using AOP in the embedded software development. In this work we propose to do embedded application development based on AOP. We implement cockpit display system application software framework (CASF) on VxWorks platform with AOP features based on AspeCt-oriented C. Based on this, we implement a sample cockpit display application and test it on the real avionic device. This case study illustrates the potential benefits and issues associated with the use of AOP in embedded development.

Keywords: embedded development, aspect-oriented programming, avionics, cockpit display application.

1 Introduction

Recently, aspect-oriented programming (AOP) [1] has been under a growing spotlight in both industrial and academic areas. AOP is a programming paradigm that aims to increase modularity by allowing the encapsulation of cross-cutting concerns. AspectJ [2] is an example of AOP implementation by language extension. It adds to Java language new constructs to express AOP concepts such as join point, pointcut, advice, inter-type declaration, and aspect. Based on JVM, AspectJ can realize both static and dynamic weaving. AspeCt-oriented C [3] is another example which offers similar AOP concepts. But due to the lack of runtime features of C language, it is implemented as a source-to-source translator which takes in C files together with aspect files and outputs normal C files. There are many other implementations of AOP [4-6]. Based on these, promising research projects have been conducted and real enterprise software have been developed (see [7]) which proved that AOP leads to more understandable, maintainable and extensible software.

Along with the benefits, AOP also raises severe concern whether it is universally applicable to all kinds of software development. One major challenge is how to apply it to the embedded software development. Current research focuses on its use in the

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_53, © Springer-Verlag Berlin Heidelberg 2014

specific software such as operating system [8], but little work has been proposed for its use in general embedded software.

In this paper, we present a case study applying AOP in the embedded software development. We have implemented cockpit display system application software framework (CASF) on VxWorks platform with AOP features based on AspeCt-oriented C. Based on this, we have implemented a sample cockpit display application and tested it on the real avionic device. We use this case study to illustrate the potential benefits and issues associated with the use of AOP in embedded development.

The rest of this paper is organized as follows. Next section presents AOP with AspeCt-oriented C. In section 3, we give a short description of CASF and how we implement it. We evaluate our work and discuss advantages and issues in section 4. Finally, we conclude this paper in section 5.

2 AOP with AspeCt-oriented C

AspeCt-oriented C (ACC) is an aspect-oriented extension of C language implemented by MSRG in University of Toronto. Unlike AspectJ, it is implemented as a source-to-source translator which outputs normal C files. This openness is especially suitable for code inspection and verification in the embedded development of safety critical systems which may need some type of certification like DO-178C [9].

When writing aspect-oriented code, four concepts are always encountered: *join point model* which defines all identifiable points in the program, *pointcut* matching a group of these points, *advice* to insert execution codes to join points matched by the pointcut, and *inter-type declaration* to append new members to existing data types.

ACC's join point model is primitive and supports four types of join points: *call join point* is the point where a function is called; *execution join point* is where a function is executing; and *set / get join point* is where a variable is assigned / read. Currently it only supports set / get join points involving global variables with base types.

For pointcuts provided by ACC, there are also four types: *primitive pointcut* which is a group of 9 basic pointcuts such as one to match call join points, *composite pointcut* which is the result of pointcut(s) doing union/intersection/complement operations, *named pointcut* aliasing existing pointcut, and *cflow pointcut* picking join points in dynamic execution context.

Similar to other aspect-oriented languages, ACC supports *before*, *after* and *around* advices with special control functions and reflective information. For inter-type declaration, ACC provides a means to add new members to existing *struct*s and *union*s by *intype pointcut* and *introduce advice*. Below is a brief code example of ACC:

```
int around(): call(int foo(int)){
  printf("around calling foo\n"); return 0;
}
```

This is an around advice indicating that every calling to function "foo" is replaced by printing out a message and setting "0" as foo's return value.

3 CASF

3.1 Cockpit Display System and CASF

The cockpit display system (CDS) is part of the cockpit avionics which provides visible human-machine interface for aircrew to manage the aircraft. Fig. 1 roughly depicts a typical CDS hardware architecture which consists of a CDS computer, some displays, some control devices, and interconnecting buses.

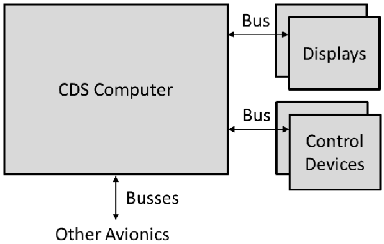


Fig. 1. A typical cockpit display system's hardware architecture

In recent years glass cockpit [10] with CDS has become widely used in both large and small aircrafts. CASF is an application software framework we propose for the development of CDS in this process. It aims to serve as a core asset base [11] to manage the commonalities and variations of different models of CDS software and to accelerate the development process of new models.

3.2 Software Architecture

In this section we present the software architecture of CASF. CASF follows the programming model suggested for general VxWorks development that there is a single application entry which spawns some concurrent tasks driven by clock interrupt. Taking advantage of AOP, CASF is more modularized as shown in Fig. 2.

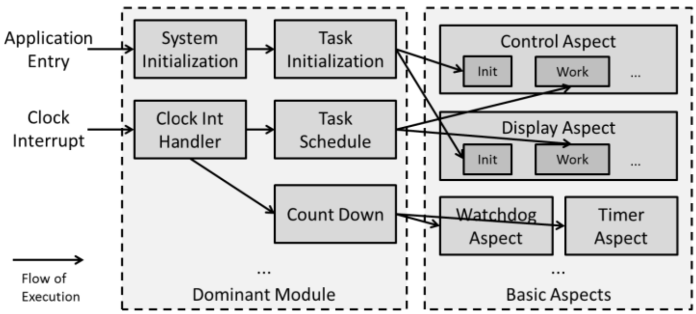


Fig. 2. Software architecture of CASF

CASF consists of two big parts: one is the dominant module which is the procedural foundation of the framework; another is a group of basic aspects realizing orthogonal concerns required by the specific application.

The dominant module mainly does two things: providing necessary abstraction of the platform's APIs and defining appropriate application model for the basic aspects to hook in. As shown in Fig. 2, it encapsulates how a task is initialized at the application entry and scheduled repeatedly by clock interrupt.

There are a handful of basic aspects, including control aspect which deals with events and state changes of the application like a button press, display aspect which routinely update the displays' screen, watchdog aspect telling that the application is still running, and timer aspect managing application defined timer events. In Fig. 2, some segments of control and display aspect are especially shown for they both involve of a task which needs initialization and repeated execution.

So, based on AOP, CASF proposes a layered architecture for the development of CDS software as depicted in Fig. 3. Upper layer software unidirectionally depends on lower layers. In this way, different layers and orthogonal aspects can be developed separately and the implementations of different concerns can be easily replaced and configured as a whole for different models of the final product.

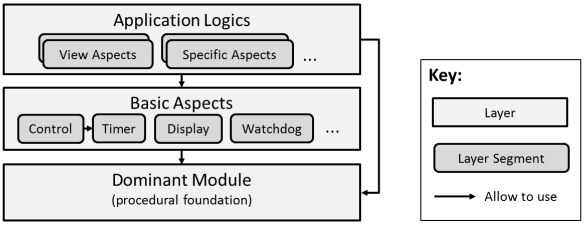


Fig. 3. Layered architecture of CDS software proposed by CASF

3.3 Design of Dominant Module

In this section we explain the design of dominant module (DM) in CASF, which is the ground work defining several fundamental constructs in CASF. The most important constructs among these are task and view hierarchy.

Task in DM represents a piece of work which has to be done periodically. It has two types: soft task and hard task. Soft task relies on the VxWorks task and is scheduled by a semaphore periodically released in the clock interrupt handler. Instead, hard task uses a countdown mechanism directly done in the clock interrupt handler. Both type of task has to specify an initialization function and a periodic work (called delegate) function which is managed by the DM in a task table. The table is originally empty and has to be filled by different aspects before the task initialization process.

As for the view hierarchy, CASF adopts the Model-View-Controller (MVC) [12] practice to structure its user interface logics. DM defines the base type and methods of view and view controller. Below are their type definitions:

```
typedef struct { } SFView;
typedef struct { SFView * view; } SFViewController;
```

As shown[1], it says merely nothing except that the view controller needs to know where its view is. Later, we will explain how control and display aspects extend these types using ACC's inter-type declaration.

3.4 Design of Basic Aspects

In this section we describe the design of basic aspects in CASF. We first give a description of the control aspect, and then talk about other aspects abstractly.

As introduced, control aspect mainly deals with events and state changes of the application. It initiates a soft task acting as an event loop, and extends "SFViewController" adding event-handling constructs. Below are some of its core AOP codes:

```
before(): execution($ SFTaskInitialization()){
  SFTaskTableAdd(SFTaskTypeSoft, "ControlTask", 25, 110,
SFControlTaskInit, SFControlTaskDelegate);
}

introduce(): intype(SFViewController){
  void (* viewWillAppear)(void *this, int screenId);
  void (* viewWillDisappear)(void *this, int screenId);
  void (* receiveInputString)(void *this, int screenId,
const char *cString, SFInputStringMode mode);
  SFResponseEntryList responseEntryList;
}
```

The first group of codes is an ACC before advice adding a new soft task (called "ControlTask") to the DM's task table before its execution of "SFTaskInitialization" function. The task has an execution rate of 25Hz, priority of 110, and initialization & delegate functions as specified.

The latter is an inter-type declaration adding event-handling constructs to the DM's "SFViewController" type. It adds three function pointers which will be called respectively when the view appears / disappears form the screen or receives input string. It also adds a "responseEntryList" member for recording which response function will be called when a specific key code event occurs.

Owing to space constraints, details of how the control task works and how events are dispatched to view controller's response functions are omitted in this paper, and we abstractly describes other basic aspects.

Like control aspect, display aspect also initiates a soft task and extends DM's data types. The task will periodically update each display's screen by calling the current[2] view's update function which is extended in the aspect.

---

1 All public types, functions in CASF have "SF" as their name prefix.

2 Each screen has an independent "current view" called root view.

As for watchdog and timer aspect, they both use quick hard task because their logic is straight forward. Watchdog aspect needs to do some checking and flip-flop a targeted memory location constantly with Interrupt handler's high execution priority, while timer aspect simply needs to manage and periodically count down timers' ticks.

Other basic aspects also exist, such as a hardware initialization aspect to preset the hardware states before the DM's system initialization, but we skip them in this paper.

4 Evaluation and Discussion

In this section we present the evaluation of CASF and discuss the advantages and issues of embedded software development based on AOP.

We implement a sample cockpit display application based on CASF to evaluate the quality of CASF. The application uses the four basic aspects described before and consists of 3 views: a main menu, a waypoint management view and a flight plan management view. Each view is implemented as an independent aspect which responses to button press events from displays and string inputs from control devices. The application also has a specific aspect to deal with the loading and saving of waypoint data in an internal NVRAM storage device.

The result is a highly modularized embedded application. Because each aspect unidirectionally depends on one another without loop, we can quickly add / remove a feature of the application by simply enabling / disabling the specific aspect and aspects depending on it. We test the sample application and all its variations (for example, with and without flight plan management feature) on the real avionic device and they work well.

From our design and evaluation result, we summarize some advantages and issues developing embedded software based on AOP. We list the items below.

4.1 Benefits

Modularity
The most important advantage provided by AOP is that it increases software modularity which is the source of understandability and maintainability. It separates different concerns of the embedded application and in our example leads to simple layered software architecture, which seldom occurs in the traditional embedded development.

Extensibility
Aspect is a high level programming abstraction which natively supports extension. In our case, the CDS software based on CASF can be easily extended by adding more basic aspects and application aspects to realize complex real-world requirements.

Configurability
Configurability means that the software can be easily configured to different models of the product with distinct features. This has been done well in our case because the aspects can be easily replaced and configured as a whole for different concerns.

Testability and Simulation

Although not addressed in this paper, AOP provides us with an easier way to do unit tests and PC simulation (by generating stub codes and hijacking platform dependent aspects) in the embedded development.

4.2 Issues

Lack of Tool Support

When we start to do embedded development based on AOP, the biggest issue is that currently there isn't much tool support available, including editor and debugger. If the editor can assist showing all applied advices to a function or type visually, it can further contribute to the understandability of the software and may eliminate some confusion.

Still Needs Good Architecture

Apparently, AOP is just a technique and it's still good software architecture that really matters. In our experience, unrestricted use of aspects promotes spaghetti code and ad-hoc design which may further complicate some problems.

5 Conclusions

This paper presents a case study applying AOP in the embedded software development. We implement cockpit display system application software framework (CASF) on VxWorks platform with AOP features based on AspeCt-oriented C. We also implement a sample cockpit display application and test it on the real avionic device. From the design and evaluation result, we summarize some advantages and issues developing embedded software based on AOP. The summary could help and inspire other embedded developers and researchers to try taking advantage of AOP in their work.

Acknowledgements. The authors would like to thank the colleagues from Software Department of China Aeronautical Radio Electronics Research Institute for their help. We furthermore thank Ming Li, Jiang Tuo for their invaluable technical assistance.

References

1. Kiczales, G., Lamping, J., Mendhekar, A., et al.: Aspect-Oriented Programming. In: Akşit, M., Matsuoka, S. (eds.) ECOOP 1997. LNCS, vol. 1241, pp. 220–242. Springer, Heidelberg (1997)
2. Kiczales, G., Hilsdale, E., Hugunin, J., Kersten, M., Palm, J., Griswold, W.G.: An Overview of AspectJ. In: Knudsen, J.L. (ed.) ECOOP 2001. LNCS, vol. 2072, pp. 327–353. Springer, Heidelberg (2001)
3. Gong, W., Jacobsen, H.A.: AspeCt-oriented C Language Specification Version 0.9. Working Technical Draft (2010), `http://www.aspectc.net`

4. Spinczyk, O., Gal, A., Schröder-Preikschat, W.: AspectC++: An aspect-oriented extension to the C++ programming language. In: Proceedings of the Fortieth International Conference on Tools Pacific: Objects for Internet, Mobile and Embedded Applications, pp. 53–60. Australian Computer Society, Inc., Darlinghurst (2002)
5. Hirschfeld, R.: Aspect-Oriented Programming with AspectS. In: Aksit, M., Mezini, M. (eds.) Proceedings of Net.Object Days 2002, Erfurt, German. (2002)
6. Spring Framework, `http://projects.spring.io/spring-framework/`
7. Aspect-Oriented Software Development Conference Home Page, `http://www.aosd.net/`
8. Coady, Y., Kiczales, G., Feeley, M., et al.: Using AspectC to improve the modularity of path-specific customization in operating system code. In: Proceedings of the 8th European Software Engineering Conference Held Jointly with 9th ACM SIGSOFT Symposium on Foundations of Software Engineering, pp. 88–98. ACM Press, New York (2001)
9. Software Considerations in Airborne Systems and Equipment Certification. RTCA, Inc., `http://www.rtca.org/store_list.asp`
10. Sweet, W.: The Glass Cockpit. IEEE Spectrum Magazine 32(9), 30–38 (1995)
11. Bass, L., Clements, P., Kazman, R.: Software Architecture in Practice, 3rd edn. Pearson Education, Inc. (2013)
12. Burbeck, S.: Applications Programming in Smalltalk-80(TM): How to use Model-View-Controller (MVC), `http://st-www.cs.illinois.edu/users/smarch/st-docs/mvc.html`

A Self-adaptive Power Management Method Based on Heartbeats

Cangzhou Yuan[1], Jiewei Ke[1], Yuxiao Liu[1], You Yue[1], and Lei Wang[2]

[1] School of Software Engineering, Beihang University, Beijing, China
yuancz@buaa.edu.cn,
{kejiewei,liuyuxia,yueyou9170}@sse.buaa.edu.cn
[2] School of Computer Science and Engineering, Beihang University, Beijing, China
wanglei@buaa.edu.cn

Abstract. For energy saving in embedded systems, numerous strategies based on CPU utilization save energy by reducing CPU operating frequency. While they're only concerned with the overall usage of system hardware, individual task's performance requirements cannot be considered separately. To improve these deficiencies, we design an energy-saving method to scale CPU performance states based on individual task performance. MIT's Heartbeats framework is applied, and improved to adapt to the characteristics of periodic tasks. A task's desired P-State can be determined by gathering real-time heartbeats and comparing them with its performance need. According to multiple tasks' P-State needs, we design two strategies to find the optimal performance state. Finally, to verify our method's effectiveness and feasibility, long-time experiments in Intel Core2 CPU and VxWorks platform are carried out. The results indicate our solution save more energy than Timeout and PAST, 2.5% and 1.7% respectively, and the performance need of each task can be satisfied.

Keywords: Adaptive system, Self-awareness, Power management.

1 Introduction

Since CPU in high load accounts for most of energy consumption, DVFS (Dynamic Voltage and Frequency Scaling) becomes the basis for many power management strategies. It can shift a chip's running fre-quency and voltage dynamically.

There are two mainstream energy-saving policies: timeout policy [1] and PAST policy [2]. Both of them are only concerned with the overall usage of system hardware, individual task's performance requirements may not be satisfied.

To improve these deficiencies, we design an energy-saving method to scale CPU performance states based on individual task performance awareness. Traditional performance evaluation methods, such as detect-ing the number of instructions executed in a period, CPU usage percent or cache hit rate, fail to realize operating process on application level. Henry Hoffmann put forward a generalized framework called Heartbeats [3] to solve application's performance self-awareness problems. With the help of Heartbeats, our method takes much consideration in each task's performance

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_54, © Springer-Verlag Berlin Heidelberg 2014

requirement which is omitted by tradition strategies and doesn't need to know task distribution feature, which means our method is applicable when new applications are added into system.

2 Related Research

The power management on system level has many research results. The basic idea of timeout policy is that when the cumulative idle time of the system exceeds a set threshold, CPU will enter the lower performance state. When there are tasks ready to run, CPU turns back into higher performance state to ensure tasks' normal running. The timeout thresh-old is critical to the algorithm. It determines the time to switch P-States, which has a great impact on energy consumption.

Prediction-based policy assumes the time to access system components is continuous, and uses the historical data to estimate the performance requirements in the future before scaling frequency. Typical examples include PAST and FLAT [4]. These policies use the overall system running status as standard of power management, so neither can meet performance requirements on task level.

Stochastic optimal policy [5] establishes a Markov model for the system pulse load, and then derives the optimal performance state based on the knowledge of state transition probabilities. The existence of optimal state has been proved and it cloud be solved in polynomial time. How-ever, this policy requires prior knowledge of task distribution and sys-tem load model, which limits its generality.

In addition, to meet the needs of real-time tasks there are some methods to improve the scheduling algorithms, such as RT-DVS [6] and SRT-DVFS. The former one scales P-States according to task schedulable conditions. The latter one takes delay threshold as basic to make frequency modification. However, they both require prior knowledge of the execution period of tasks, and the time varies on different machines.

3 Power Management Method Based on Heartbeats

The critical problem of power management is the conflict between energy efficiency and performance. To take into account the performance needs of individual task, we propose a solution to make each task run-ning in its desired performance state, shown in Figure 1.

Our solution requires the developers to follow Heartbeats framework specification: insert Heartbeat API function in the end of the critical loop of application (usually select the outermost loop to reduce the overhead caused by Heartbeats), and set a desired heart rate range for each task. Calculating the time difference between two successive heartbeats and taking its reciprocal can obtain task running rate.

Power management monitor reads task heart rate records at regular time, and judges whether each task can meet performance requirements by comparing it to heart rate range. If not, it will select an optimum CPU performance state P_{cur}, access the CPU drives to manipulate DVFS control interface, and switch current P-State to a higher or lower one.

Two key issues will be discussed in more detail below, including how to perceive the performance of various tasks and find their desired P-States; how to choose an optimal system performance state according to multi-tasks' needs.

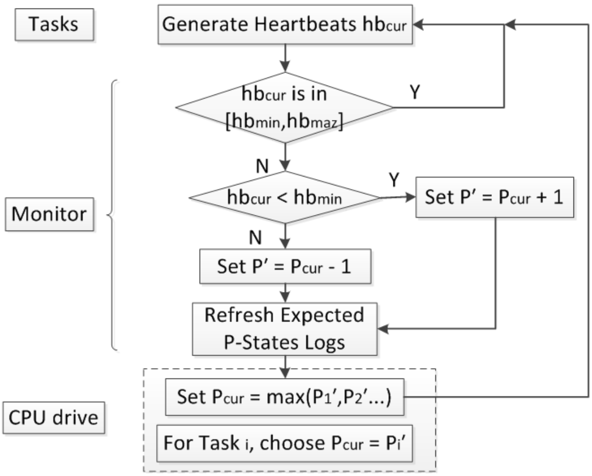


Fig. 1. Self-adaptive power management schematic diagram

3.1 Performance Awareness

Here is an experiment of scaling CPU P-States during a single task running, and the volatility of task heart rate is shown in Figure 2. The experimental CPU has five P-States, and we gradually scale it from P_4 to P_0. At each P-State, the task heart rate remains stable, reflecting the periodic task operation status. After P-State scaling, the task heart rate will immediately occur obvious transition, and in a very short time step into a new narrow range. That means, Heartbeats framework can be perceived as a self-awareness tool for power management.

However, how to identify each task's desired P-State? If heart rate is outside the expected range, we scale current P-State to the nearby one, and then verify if the new heart rate meets the expected range. Usually after a few attempts (the worst case is the number of P-States supported by CPU subtracts one, in our experiment it is 4 times), we get the task's desired P-State. In Figure 2, assuming that the expected heart rate range of task T_1 is [6.7,6.9], we find that P_1 is the expected P-State of T_1. There is a situation that if desired heart rate range is set too narrow, P-State jolts will happen and the heart rate can never get into the range. It can be solved by inserting redundant information to record P-States scaling history, and P-State can be fixed at the higher one of the two fluctuant states, which will meet task performance requirements.

As for how to set heart rate range, we know that heart rate is related to procedures operation scale. Since most embedded developers are experienced, and lots of tools such as WindRiver WindView, can help identify program operation time, the work of setting heart rate range is not difficult.

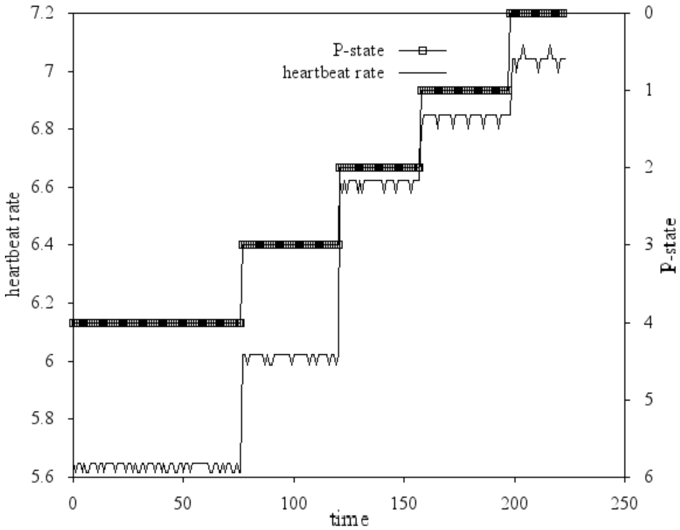


Fig. 2. At each P-State, the task heart rate is stable; and when P-State switches, heart rate is sensitive to reflect changes in execution speed

3.2 Optimal System Performance State

The basic idea of our method is to regularly monitor current heart rate of each task, and make a decision whether to adjust CPU P-States. Suppose we know all tasks' desired P-State, which forms a P-State set P_1', P_2' ..., we have two strategies.

Strategy 1: Adjust the CPU to the highest P-State of the set, which ensures all tasks to meet performance requirements.

$$P = \max(P_1', P_2' \ldots) \tag{1}$$

However, since just one task can gain CPU each time, the strategy above attempts to satisfy performance requirements of all tasks and causes some waste. So we propose an improved strategy, where actual system P-States adjustment does not occur during monitoring, but in the process of task switching.

Strategy 2: During monitoring process, compare the previous heart rate with the desired range, if P-State scaling is needed, then update the desired P-State records. When a task gets CPU, we scale system P-State in accordance with its desired P-State record.

$$\text{For Task i,choose} \quad P = P_i' \tag{2}$$

The remaining question is whether a new arriving task will cause system instability. We have accounting information record for each task's desired P-State, in this way, the arrival of new tasks will not interfere with the operation of each task when obtaining CPU resources. In ac-cordance with the original record information, system P-State can be adjusted to the desired one in a step, or only a few adjustments require.

4 Experiments and Results

To verify energy-saving effectiveness of our method, we carried out simple implementation of timeout and PAST for contrast. We employ PC with Intel Core2 E7400 processor as target machine, which supports Intel's SpeedStep technology. And VxWorks is chosen as our experimental target system, as kernel mode and user mode do not distinguish in VxWorks, which simplifies implementation to access registers related to power management. JNX2000D, a wireless power metering outlet, is used to measure energy consumption.

System generates a random sequence of 200 tasks, and each task's start time and total workload are also randomized. Such programs can simu-late the scenario that tasks with different lengths co-exist in real-time system. Similar simulation is applied for the real-time embedded microkernel assessment[7] and RT-DVS energy measurements.

With the same tasks sequence running in VxWorks, we measured energy consumption for four cases: without power policy, timeout policy, PAST policy and our second strategy. And 20 groups of experiments were carried out with diverse task sets. Results show in Table 1.

Table 1. Characteristic and energy consumption of different policies

Policy	Basic for scaling	Overhead	Avg Energy(Joule)
Without policy	None	None	257920
Timeout	Length of idle time	None	257040
PAST	CPU utilization rate	None	255200
Our policy	Current heart rate and the input heart rate range	Generate heartbeats	250930

Timeout policy's energy-saving effect is not obvious, because it scales P-State to the highest state P_0 to reduce task delay when a task begins to run. Waiting before timeout also brings a certain energy waste. PAST considers general system performance requirements for all tasks, while most of the tasks may not need such a high run-time P-State, so it wastes some energy too.

Calculating from Table 2, our power management method based on Heartbeats saves 2.7% energy. The test is characterized by quite frequent task switching, and the task switch hook function brings a large overhead. We have also designed a test case with less task switching for longer-time measuring, and find that its energy saving efficiency rise to 5.1%. If not considering energy consumption by memory, hard drives and other components, the benefit will be greater. However, by measuring 20 times and taking energy consumption of other components simplified as a constant, it sufficiently proves that compared with the other two policies, our proposed method has better energy saving effect for systems where tasks with different performance requirements co-exist.

During the experimental program, two tasks are extracted to analyze our energy policy's effect on task performance. When a task obtains CPU, P-States can be scaled based on its operation status without considering the impact of other tasks. It ensures efficiently energy usage for a uniprocessor system. By building and solving the equations with instructions number, CPU frequency and executing time of P-State transition, we get the latency for CPU to reach the target state is 0.0016s, which can be ignored. However, if task switching is frequent in system, it will cause a larger sum of overhead. Therefore, we should choose a proper strategy according to the application scenarios in practice.

5 Conclusions

This paper proposes a self-adaptive power management method for multi-task system, with Heartbeats framework as basis for performance evaluation. Verified by experiments , we find that our method can mitigate the status quo of increasing energy consumption and ensure multi-task performance requirements. Our future work in this area will consider scheduling factors to solve tasks mutual preemption problems.

References

1. Srivastava, M.B., Chandrakasan, A.P., Brodersen, R.W.: Predictive system shutdown and other architectural techniques for energy efficient programmable computation. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 4(1), 42–55 (1996)
2. Weiser, M., Welch, B., Demers, A., Shenker, S.: Scheduling for reduced CPU energy. In: Mobile Computing, pp. 449–471. Springer (1996)
3. Hoffmann, H., Eastep, J., Santambrogio, M.D., Miller, J.E., Agarwal, A.: Application heartbeats for software performance and health. ACM Sigplan Notices 45, 347–348 (2010)
4. Pering, T., Burd, T., Brodersen, R.: The simulation and evaluation of dynamic voltage scaling algorithms. In: Proceedings of the International Symposium on Low Power Electronics and Design, pp. 76–81 (1998)
5. Benini, L., Bogliolo, A., Paleologo, G.A., De Micheli, G.: Policy optimization for dynamic power management. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 18(6), 813–833 (1999)
6. Pillai, P., Shin, K.G.: Real-time dynamic voltage scaling for low-power embedded operating systems. ACM SIGOPS Operating Systems Review 35, 89–102 (2001)
7. Zuberi, K.M., Pillai, P., Shin, K.G.: Emeralds: A small-memory real-time microkernel. In: Sixth International Conference on Real-Time Computing Systems and Applications, RTCSA 1999 (1999)

Schedulability of Real-Time Systems with Enhanced Safety

Dingkun Yang and Fei Hu

School of Software, Shanghai Jiaotong University, Shanghai, P.R. China
ydkwhitebear@gmail.com, hufei@sjtu.edu.cn

Abstract. A real-time system consists of various tasks and dependencies between these tasks. Due to the time sensitive nature of such tasks, both time constraints and dependencies should be satisfied. To ensure each task running correctly and completely, especially critical tasks that affect the safety of system, some tasks will have to be missed, a reasonable algorithm is needed to find a compromise between task dependencies and system safety. In order to illustrate the algorithm, a typical real-time system - small unmanned helicopter onboard software system is introduced as an example. Through discussing the task model of onboard software and the task dependencies sequence, the problems caused by task dependencies is shown, a new task model and an algorithm based on the new model to solve above problems is discussed and proposed. Experiments show the improved results, finally is the further discussion on the algorithm.

Keywords: real-time task, scheduling, UAV software, dependency, safety.

1 Introduction

A real-time system can be loosely defined as a system whose response time is an important determinant of correct functioning[1]. A real-time system consists of various tasks and dependencies between these tasks. Due to the time sensitive nature of such tasks, both time constraints and dependencies should be met, and if not, the whole systems may go wrong, or even produce serious consequences. Therefore, to ensure each task running correctly and completely, especially critical tasks that affect the safety of system, a reasonable algorithm is needed to find a compromise between task dependencies and system safety.

In order to illustrate task dependencies and system safety, system modeling is needed. A typical real-time system - small unmanned helicopter onboard software system is introduced as an example. Such software systems are required to perform tasks from hardware driving to the management of device operation, and from traditional input-output control law implementation to task scheduling and event disposal[2]. Onboard software system is the core component of the small unmanned helicopter, which consists of the flight control, communication, aviation task application, emergency processing and other components.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_55, © Springer-Verlag Berlin Heidelberg 2014

There have been many researches on onboard system modeling and real-time schedulability that use unmanned helicopter software systems as examples. Liu et al. took small unmanned helicopter as an example of real-time multi-tasking system scheduling[3]. Sagahyroon et al. discussed how to optimize UAV controller design, modeling, implementing and testing to make the controller overhead smaller[4]. Pastor et al. proposed a special hardware/software architecture[5]. Santamara et al. analyzed unmanned flight system software architecture, and mainly focused on the real-time software development process[6]. Oreizy et al. presented self-adaptive software applied on UAV software systems[7]. Mutter et al. introduced a virtual test environment for UAV software are the modeling and simulation of the environment and the hardware platform, as well as the integration of the software in the simulation[8]. Tang et al. proposed a reconfigurable real-time onboard control system for a small UAV helicopter[9]. Valenti et al. discussed the implementation of UAV mission system scheduling[10].

Based on above researches, this paper focuses on real-time tasks in onboard software system which are periodic, non-preemptive and indivisible. The paper is organized as follows: Section 2 defines the task model and discusses the task dependencies sequence. Section 3 shows the problems caused by task dependencies. Section 4 proposes and discusses a new task model and proposes a scheduling algorithm based on the new model to solve above problems. Section 5 shows the improved results. Section 6 further discusses the algorithm.

Table 1. Onboard software task model

T_{no}	TaskName	$T_a(ms)$	$T_e(ms)$	$T_p(ms)$	$T_d(ms)$	Deps
1	MsgRetrieving	0	4	50	50	
2	MsgDataAnalysis	0	2	50	50	1
3	GyroDataRetrieving	0	2	20	20	
4	ComDataRetrieving	0	3	20	20	
5	UsaDataRetrieving	0	1	100	100	
6	GPSDataRetriving	0	2	100	100	
7	InsAnalysis	0	2	20	20	2
8	SnrDataAnalysis	0	2	20	20	3,4,5,6
9	CtrlLawCalculation	0	2	20	20	7,8
10	PWMGeneration	0	1	20	20	9
11	DataStorage	0	10	200	200	9
12	MsgEncapsulation	0	2	50	50	11
13	MsgSending	0	3	50	50	12

2 Software Task Model

Tasks discussed in this paper are periodic, non-preemptive and indivisible. Each task is defined as a 7-element tuple: (T_{no}, *TaskName*, T_a, T_e, T_p, T_d, *Deps*). T_{no} is the only identity of the task. *TaskName* is task name. T_a is task arriving time, T_e is task execution time. T_d is task deadline, a task must finish its execution before deadline or it will be missed. T_p is task period. *Deps* contains the T_{no} s of the task dependencies, all dependencies should form a topological sequence. T_{no} is a positive integer, *TaskName* is a string, T_a, T_e, T_p, T_d are non-negative integers and *Deps* is a positive integer array. According to the task definition, a 13-task onboard software model is established, which is based on the RTIO lab helicopter software system in Shanghai Jiao Tong University [11]. Each task describes in Table 1.

Task 1 receives data from ground station and puts it into buffer. Task 2 gets data from the buffer and decodes it into instructions. Task 7 further decodes the instruction for task 9. Tasks 3,4,5,6 are the 4 main sensors that collect position data of the helicopter. Task 8 decodes data from 4 different sensors and sends the data to task 9. Task 9 calculates and generates the aircraft control data. Task 10 generates the PWM signals for servo control according to the result from task 9. Tasks 11,12,13 are in responsible for flight data receiving, storing and sending, in convenience for ground station monitoring.

As no loopback is allowed, tasks and their dependencies can form a **Directed Acyclic Graph(DAG)**. Then, how to define dependencies between tasks with different periods and arriving times need to be declared. Without affecting the dependencies, here we calculate **Least Common Multiple(LCM)** of all tasks' periods, and name it **Compound Period(CP)**. In a **CP**, all tasks will be executed at least once. Another concept is **Shortest Period(SP)**, which refers to the shortest period in all. Below discussions are based on **CP** and **SP**. In our discussion, **CP** is 200 ms and **SP** is 20 ms, there are 10 **SP**s in the **CP**. In every **SP**, to generate a task priority sequence, both topological sequence and task scheduling algorithm should be considered. Here we proposes the EDF Topological Sequence S_{et}. Below is the algorithm for achieving S_{et}.

① If the task DAG is not empty, calculate all nodes' in-degree ID_is,

② Scan all $ID_i = 0$ nodes, use **Early Deadline First(EDF)** find the first task to be executed, add this node into Set, delete it and its post relationship in the task **DAG**, update the relationships; if a task arrives, add both its pre-and-post relationships, update the relationships; if a task reaches its deadline without being executed, delete this task node and its pre-and-post relationship in the DAG, update the relationships, return to ①.

For instance, use the algorithm, according to the tasks in Table 1, in time 0 ms, task 1 to 13 has arrived, Set is 3, 4, 1, 2, 7, 5, 6, 8, 9, 10, 11, 12, 13; in time 20 ms, task 3, 4, 7, 8, 9, 10 has arrived, task 10 is missed for it has reach the deadline without being executed, Set is 3, 4, 7, 8, 9, 10, 11, 12, 13.

3 Problems

However, S_{et} can still cause task missing. Below list some of the problems.

3.1 Problem 1: Task Missing Caused by Task Dependencies

In a CP, tasks may delay because they have to wait for their dependency tasks, even though the total running time of tasks in CP may be less than CP length.

Just take the tasks in Table 1 for example, in CP, task 3,4,7,8,9,10 will be executed 10 times, task 1,2,12,13 four times, task 5,6 twice and task 11 only once, their total execution time is 180 ms, which is less than CP. However, at 20, 120(ms), task 9, 10 are missed, and at 50, 100, 150, 200(ms), task 12, 13 are missed, at 200(ms), task 11 is missed. These tasks are still missed because they have to wait for their dependencies.

3.2 Problem 2: Task Missing Caused by Arriving Time Delay

In a CP, tasks may delay because of their dependencies' delay. For instance, this time, the task execution times from task 1 to task 13 change to 3, 2, 2, 3, 1, 2, 2, 2, 2, 1, 5, 2, 3(ms), their total execution time within a CP is 171 ms, and use Set, they will be executed without missing any task.

But start from **SP** 3, task 4's arriving time has 15 ms delay for some reasons. Use EDF Topological Sequence, the result shows, at 60, 80, 100, 120, 140, 160, 180, 200(ms), task 9, 10 are missed, and at 100, 150, 200(ms), task 12, 13 are missed. These tasks are missed because they depend on the delayed task 4.

3.3 Problem Analysis

These problems will cause onboard task missing, even threaten the aviation safety. For instance, task 10 is missed in both 2 problems, this will cause servo failure, furthermore the control of aircraft will be lost and the aircraft may crash. We cannot make fully prediction of these 2 problems when we design the software system. But through analysis, we know that of all onboard tasks, some tasks affect the flight safety; while some others only affect the accuracy, quality or even later reproduce, but not safety.

Based on this idea, we can solve this problem by adding more attributes into the task definition, especially the attributes related to safety and dependency. That is, we can ensure tasks with higher safety and stronger dependencies in advance, thus greatly enhance the safety of onboard software.

4 Improved Task Model and Scheduling Algorithm

4.1 Improved Task Model

Based on the idea of the previous section, some concepts are proposed here.

Task Dependency Degree: task A depends on task B, the weight of dependency is called task dependency degree, denoted by $D_{A,B}$. Task dependency degree is a positive integer, higher the value is, A is more dependent on B. D_φ is the threshold of task dependency degree, if $D_{A,B} > D_\varphi$, the dependency between A and B is strong dependency; otherwise is weak dependency.

Safety Degree: the importance of the task in the whole software system, denoted by S_i. Safety degree is a positive integer, higher the value is, the task is more crucial to the whole system. S_φ is the safety degree threshold, if the safety degree $S_i > S_\varphi$, then the task is a high safety task, otherwise is a low safety task.

Now each task is defined as a 9-element tuple: (T_{no}, *TaskName*, T_a, T_e, T_p, T_d, *Deps*, *DepDegrees*, *Safety*). *DepDegrees* is a positive integer array, and *Safety* is a positive integer. In our system, $D_{2,1}$=1, $D_{7,2}$=1, $D_{8,3}$=3, $D_{8,4}$=2, $D_{8,5}$=1, $D_{8,6}$=3, $D_{9,7}$=2, $D_{9,8}$=2, $D_{10,9}$=3, $D_{11,9}$=1, $D_{12,11}$=1, $D_{13,12}$=1 and D_φ=2. The safety degrees from task 1 to task 13 are 2, 2, 3, 2, 2, 1, 2, 3, 3, 3, 1, 1, 1. S_φ=2. So task 1 to 5 and 7 to 10 are high safety tasks, the rest are low safety tasks.

Average Importance Degree: For task t_i, OD_i is t_i's out degree in the task DAG, Ts_i is t_i's direct successor task set, for each t_j in Ts_i, we have t_j's safety degree S_j and its dependency degree $D_{j,i}$. The **Average Importance Degree(AID)** is defined in formula (1).

$$\mathrm{AID} = \frac{\sum_{j=1}^{ODi}(SjDj,i)}{ODi} \tag{1}$$

Task Levels: For a given task set, because S_{et} is partial order, if any task t_i is delayed or missed, only its direct subsequent tasks will be directly influenced. So the nature of t_i's direct subsequent task set Ts_i and related relationships is important. Any task should be one of the below 6 levels according to their safety degrees and subsequent task sets:

Low safety tasks belong to level 1,2,3, high safety belongs to 4,5,6: level 1 and 4 tasks have no subsequent task or all subsequent tasks are weak dependent on them; level 2 and 5 tasks have subsequent tasks and at least one but not all subsequent tasks are strong dependent on them; level 3 and 6 tasks have subsequent tasks and all subsequent tasks are strong dependent on them.

Level 1 to 3 tasks can be delayed and missed, others can only be delayed.

4.2 Improved Scheduling Algorithm

With above discuss, an algorithm is proposed to find task to delay or miss.

① For Set in SP, if there is some high safety tasks missing, mark the first missing task as tf , set integer i as task level, i is initialized to 1 and i is global;

② Scan all the tasks in Set, if two or more tasks belong to level i, calculate their AIDs, choose the smallest; for the same AID, choose the task with longest running time. For the chosen task, if i<=3, go to④; if i>3, go to⑤; else if i<7, go to③; else, go to⑥;

③ If no task belongs to level *i*, let *i* increment 1, back to②;

④ If the chosen task in S_{et} is before t_f , put the chosen task just after tf in S_{et}; if the chosen task in S_{et} is after t_f , put the chosen task at the end of S_{et}. If schedulable, end the algorithm. If there is still high safety task missing, miss the chosen task. If there is still high safety task missing, back to②;

⑤ If total execution time is shorter than CP length, put the chosen task at the end of Set. If this time is schedulable, end the algorithm. If there is still high safety task missing, back to②;

⑥ No available sequence, end the algorithm.

5 Test and Result

Here applying the improved scheduling algorithm in section 4 to the problems in Section 3.

5.1 Improved Experiments for Problem 1

Fig. 1 shows the relevant task miss rate and high safety miss rate, miss rate here is defined as the quotient of the number of tasks missed in SP and the number of tasks, and high safety task miss rate here is defined as the quotient of the number of high safety tasks missed in SP and the number of tasks.

As shown in Fig. 1, after improvement, the task miss rate in CP has reduced to zero in most time, only 1 task missed, and the high safety task miss rate is always zero, the system is much safer.

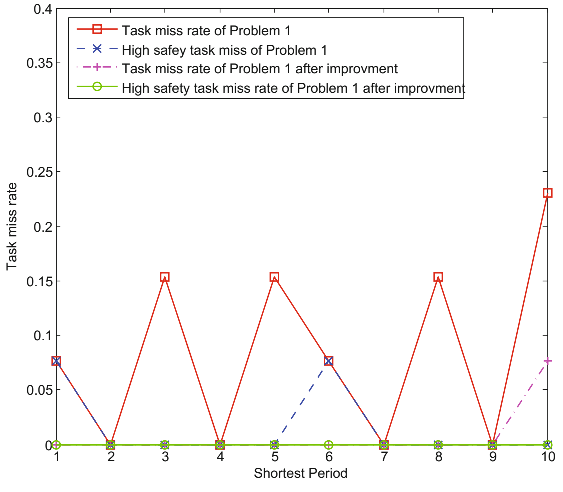


Fig. 1. Miss rate of Problem 1 before and after improvement algorithm

5.2 Improved Experiments for Problem 2

And the relevant task miss rate and high safety miss rate shows in Fig. 2. As shown in Fig. 2, after improvement, the task miss rate in CP has greatly reduced, and the high safety task miss rate is always zero, system is much safer.

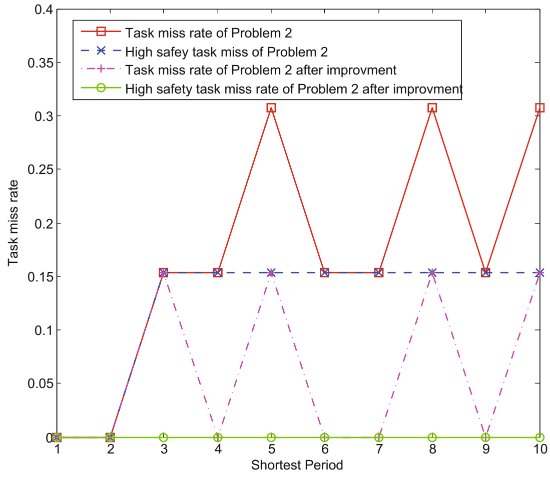


Fig. 2. Miss rate of Problem 2 before and after improvement algorithm

6 Further Discussion

There is still a case that all high safety tasks are not schedulable in a CP. For instance, the task execution times from task 1 to task 13 change to 4, 2, 2, 3, 5, 2, 4, 4, 4, 3, 10, 2, 3(ms), the high safety task execution time in a CP is 229 ms, even miss all the low safety tasks, the system is still not schedulable. Under this condition, at least one high safety task must be missed, although they are crucial to the system. As a choice, we can try to reduce one of the high safety task's level to low safety, list its miss rate and

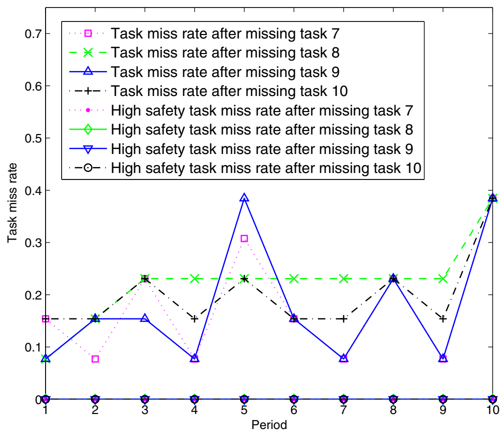


Fig. 3. Further discussion task miss rate

let the user do the comparison. Fig. 3 shows the task miss rate if we miss task 7, 8, 9 or 10 separately, and in each of the four conditions, high safety task miss rate has reduced to 0. This time, user will have to choose one from the 4 tasks and miss it.

7 Conclusion

This paper focuses on the conflict between task dependencies and system safety, uses the onboard software system as example, illustrates and discusses the task model, proposes an algorithm to solve the problem. Experiments show task miss rates especially high safety miss rate are greatly descended by using the algrithm, the safety of system can be enhanced. For further discussion, in addition to move the judgment to user, future work will focus on multicore scheduling.

References

1. Krishna, C.M.: Real-Time Systems. McGraw-Hill Education, India (1997)
2. Miaobo, D., Ben, M.C., Guowei, C., Kemao, P.: Development of a Real-time Onboard and Ground Station Software System for a UAV Helicopter. Journal of Aerospace Computing, Information and Communication 4, 933–955 (2007)
3. Jie, L., Edward, L.: Timed Multitasking for Real-Time Embedded Software. IEEE Control Systems Magazine 23, 65–75 (2003)
4. Sagahyroon, A., Jarah, M.A., A1-Ali, A., Hadi, M.: Design and Implementation of a Low Cost UAV Controller. In: IEEE International Conference an Industrial Technology, pp. 1394–1397. IEEE Press, Hammamet (2004)
5. Enric, P., Juan, L., Pablo, R.: UAV Payload and Mission Control Hardware/Software Architecture. IEEE A&E Systems Magazine 22, 3–8 (2007)
6. Daniel, S., Francisco, A., Antonio, J., Antidio, V., Manuel, B., Anibal, O.: Model-Based Design, Development and Validation for UAS Critical Software. J. Intell. Robot Syst. 65, 103–114 (2012)
7. Oreizy, P., Gorlick, M.M., Taylor, R.N., Johnson, G., Medvidovic, N., Quilici, A., Rosenblum, D.S., Wolf, A.L.: An Architecture-Based Approach to Self-Adaptive Software. IEEE Intelligent Systems 14, 54–62 (1999)
8. Mutter, F., Gareis, S., Schätz, B., Bayha, A., Gruneis, F., Kanis, M., Koss, D.: Model-Driven In-the-Loop Validation: Simulation-Based Testing of UAV Software Using Virtual Environments. In: 18th Engineering of Computer Based Systems (ECBS) (2011)
9. Yi-Rui, T., Yangmin, L.: Design of an optimal flight control system with integral augmented compensator for a nonlinear UAV helicopter. In: 10th World Congress on Intelligent Control and Automation, pp. 3927–3932. IEEE Press, Beijing (2012)
10. Valenti, M., Schouwenaars, T., Kuwata, Y., Feron, E., How, J., Paunicka, J.: Implementation of a Manned Vehicle-UAV Mission System. In: AIAA Guidance, Navigation, Control Conf. Exhibit. (2004)
11. RTIO lab in Shanghai Jiaotong University, `http://rtio.se.sjtu.edu.cn`

Dual Simulation of Application Specific Logic Controllers Based on Petri Nets

Arkadiusz Bukowiec and Jacek Tkacz

University of Zielona Góra, Institute of Computer Engineering and Electronics,
ul. Licealna 9, 65-417 Zielona Góra, Poland
{a.bukowiec,j.tkacz}@iie.uz.zgora.pl

Abstract. The paper describes a new method of verification of the application specific logic controllers designed with use of Petri net. The method is oriented on behavioral simulation of post synthesis models. Two different method of logic synthesis are preformed for the same control algorithm described as Petri net. Then, obtained HDL models are used for dual simulation with result comparison. It allows to detect ambiguous construct in designed control algorithm.

Keywords: logic controller, Petri net, simulation.

1 Introduction

The paper presents a new method of dual simulation of the application specific logic controllers (ASLCs). ASLCs [1-5] are one of the biggest group of electronic circuits. The most classical way of designing such controllers is application of hardware description languages (HDLs) but it is annoying for designer and possibly it gives high risk of human mistake. The usage of graphical representation of control algorithm is much more conformable [6-8]. Now, Petri nets (PNs) [9, 10] are one of the most adequate methods for formal design of such controllers [11]. It gives a simple way for representation of concurrent processes and application of interpretation gives a possibility to exchange information with environment. Additionally, there could be applied mathematical algorithms for formal analysis and verification of the designed model [12-15]. Although these algorithms are able to detect defects, like liveness or reachability, in Petri net structure but they are not able to detect nondeterministic behavior of the algorithm. Typically the nondeterministic behavior of the algorithm is caused by a human mistake and it does not affect Petri net structure. To detect such mistake there have to be performed the simulation of the control algorithm and the comparison of its results with expected ones. It is very arduous process and it does not give expected results because it is also loaded with human mistake. The solution for this problem is dual simulation [16, 17] with automated results comparison. Although currently only behavioral HDL description is used for such approaches and it is oriented on increase of simulation reliability. It allows also to detect some nondeterministic behavior but while the simulation is on behavioral level not connected to a hardware realization it does not ensure that all of them will be detected and controller

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_56, © Springer-Verlag Berlin Heidelberg 2014

will work in the same manner in case of a hardware realization. In presented approach there is proposed to perform two logic syntheses of tested control algorithm. There should be used two different synthesis algorithms oriented on the same hardware architecture. As a results there will be received two different logic descriptions of the same ASLC. Then dual simulation can be performed. Because different synthesis methods should interpret nondeterministic behavior in different way it could be easy detected during such simulation. It is also oriented on hardware realization, so the results are reliable.

2 Petri Net Definition

A *Petri net* [9, 10] is defined as a quadruple

$$PN = (P, T, F, M_0), \tag{1}$$

where:

- P is a finite non-empty set of *places*, $P = \{p_1, \ldots, p_M\}$,
- T is a finite non-empty set of *transitions*, $T = \{t_1, \ldots, t_S\}$,
- F is a set of *arcs* (describing flow relations) from places to transitions and from transitions to places: $F \subseteq (P \times T) \cup (T \times P)$, $P \cap T = \varnothing$,
- M_0 is an initial marking.

Sets of *input* and *output* transitions of a place $p_m \in P$ are defined:

$$\bullet p_m = \{t_s \in T : (t_s, p_m) \in F\},$$
$$p_m \bullet = \{t_s \in T : (p_m, t_s) \in F\}.$$

By similarity Sets of *input* and *output* places of a transition $t_s \in T$ are defined:

$$\bullet t_s = \{p_m \in P : (p_m, t_s) \in F\},$$
$$t_s \bullet = \{p_m \in P : (t_s, p_m) \in F\}.$$

A *marking* of a Petri net is defined as a function: $M : P \to N$.

Intuitively, given a place p_m the function $M(p_m)$ returns the number of tokens in p_m. A place or a set of places is *marked* when it contains a token. A transition t_s can be *fired* if all its input places are marked. Firing a transition removes tokens from its input places and puts one token in each output place.

A Petri net is called k-bounded if any place does not contain more than k tokens in all its reachable markings, including the initial marking. The Petri net is *safe* if it is 1-bounded.

A Petri net enhanced with an feature for information exchange is called an *interpreted* Petri net [10]. This exchange is made by use of binary signals. Boolean variables occurring in the interpreted Petri net can be divided into three sets:

- X is the set of input variables, $X = \{x_1, \ldots, x_L\}$,
- Y is the set of output variables, $Y = \{y_1, \ldots, y_N\}$,
- Z is the set of internal variables (most cases not used, with $Z = \varnothing$).

An interpreted Petri net has a guard condition φ_s associated with every transition t_s. The guard condition φ_s is defined to be a Boolean function of a subset of variables from the sets X and Z. In a special case, the condition φ_s can be defined as logic 1. Now, a transition t_s can be fired if all its input places are marked and the current value of the corresponding Boolean function φ_s is equal to logic 1. The conjunction ψ_m associated with a place p_m is an elementary conjunction of positive literals formed from output variables from the set Y and Z. If the place p_m is marked, the variables from corresponding conjunction ψ_m are set to logic 1 and other variables are reset.

Only interpreted safe Petri nets can used as models of control algorithms for ASLC.

3 Dual Simulation

There is proposed to verify control algorithm by applying dual simulation procedure (Fig. 1). This procedure is based on simultaneous simulation of two models. These two models are obtained during dual synthesis process. There should be used two different synthesis algorithms oriented on the same hardware architecture. There are used modular logic synthesis [18] and architectural synthesis [19] in presented approach but other methods are also acceptable [2, 3, 20-22]. Both models are described as set of logic equations received from synthesis process. In order to perform the simulation both models are described in HDLs.

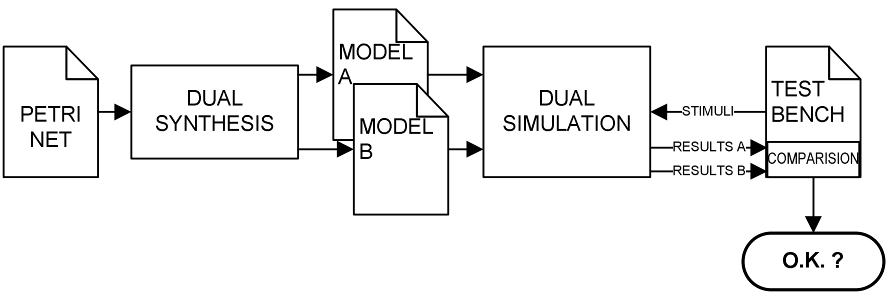


Fig. 1. Dual Simulation Flow

For the simulation, there is used a testbench. It is also written in HDL. The testbench generates the same input vectors for both models at the same time. Additionally, it compares the output signals. If the comparison process does not detect any differences it means that the control algorithm is proper design. Because there are used two different formal methods of synthesis, that are mathematically proofed, there are no possibility to obtain wrong models. If the comparison process detects any differences it means that Petri net model is nondeterministic and it have to be corrected be the designer.

The dual simulation approach is illustrated by application for sample Petri net (Fig. 2) This net has 7 places and 7 transitions. It is extended with interpretation by adding 5 input signals x_1, x_2, x_3, x_4, and x_5 and 4 output signals y_1, y_2, y_3, and y_5. The Petri net is well designed. It is safe and live.

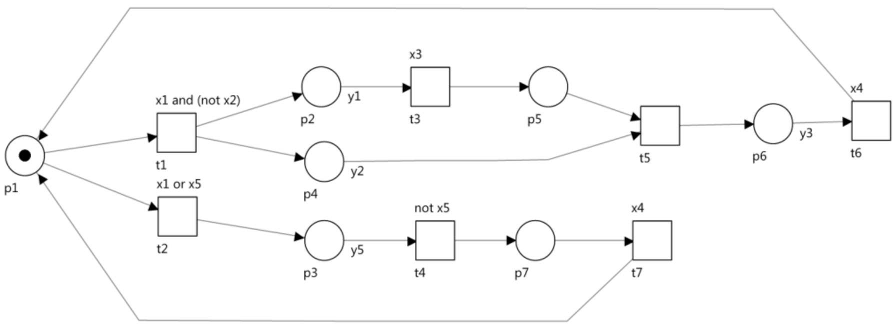


Fig. 2. Sample Petri Net

But, the place p_1 has two output transitions t_1 and t_2. It means that in view of classical definition of Petri net it is nondeterministic. After application of interpretation it could became deterministic if conditions assigned to transitions satisfy this. The presented sample net is nondeterministic because there are such conditions: $\varphi_1 = x_1 \overline{x_2}$ and $\varphi_2 = x_1 + x_5$. This is defect in case of designing an algorithm for ASLC because there is no possibility to determine how it will be executed in hardware. To detect such defect the dual simulation is applied. First, there have to be performed dual synthesis with use of two different synthesis methods [18, 19]. The obtained HDL models are instantiated in one testbench. The testbench performs also outputs comparison of both models and it generates the status signal OK. Iff values of outputs are different it is indicated by logical 0 otherwise the its value is set to logic 1.

The results of sample simulation of sample Petri net are shown in figure 3. The outputs of first model are mapped to the bus Y_a and the outputs of second model are mapped to the bus Y_b. First, the inputs x_1 and x_2 are set in 300ns. It means that only the transition t_2 can be fired and both models operate the same way. They set the output y_5 in 400ns and then they fire transitions t_4 and then t_5 and go back to the place p_1. Then, in 1100ns, there is set only the input x_1. It means that both transitions t_1 and t_2 can be fired. Now, both models behave different. First model marks only the place p_2 and it generates the output y_1, the second model marks places p_2, p_4, and p_3 and it generates outputs y_1, y_2, and y_5. This situation is detected by the testbench and indicated by reset of the status signal OK.

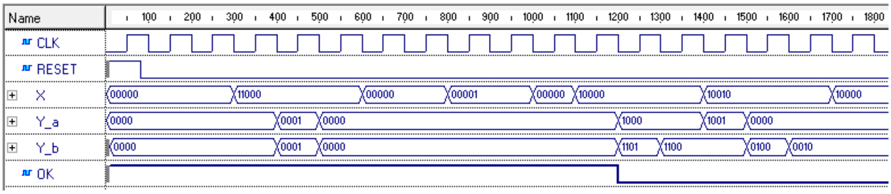


Fig. 3. Dual Simulation of Sample Petri Net

4 Summary

The paper concentrated on the verification of the specification of application specific logic controllers. We assume that the specification of control algorithm is designed as a Petri net. The mathematical algorithms are able to detect structural errors in Petri net [10, 23] but there is no simple way to detect human mistake that leads to possibility of nondeterministic interpretation of some constructs. The proposed solution is based on dual simulation of different models of the same algorithm. The only way to obtain different equivalent models is application of different synthesis methods[2, 3, 18-22]. The application of two formal synthesis methods gives to different mathematical description of the same algorithm. Now, the simulation with results comparison is able to detect nondeterministic constructions. The compression procedure can be extended to comparison of internal states, although the representation of internal states could be different after synthesis there should be included decoding of these states. The proposed verification method is easy and fast to application. It can be performed with any HDL simulator. Additionally, dual models could be use in implementation of logic controllers for safety critical systems [16, 17, 24, 25].

References

1. Chang, N., Kwon, W.H., Park, J.: Hardware implementation of real-time Petri-net-based controllers. Control Eng. Pract. 6(7), 889–895 (1998)
2. Marranghello, N., Mirkowski, J., Bilinski, K.: Synthesis of synchronous digital systems specified by Petri nets. In: Hardware Design and Petri Nets, pp. 129–150. Kluwer, Boston (2000)
3. Gomes, L., Costa, A., Barros, J., Lima, P.: From Petri net models to VHDL implementation of digital controllers. In: 33rd Annual Conf. IEEE Industrial Electronics Society, IECON 2007, Taipei, Taiwan, pp. 94–99 (2007)
4. Wiśniewski, R., Barkalov, A., Titarenko, L., Halang, W.: Design of microprogrammed controllers to be implemented in FPGAs. Int. J. Appl. Math. Comput. Sci. 21(2), 401–412 (2011)

5. Khamis, A., Zydek, D., Borowik, G., Naidu, D.S.: Control system design based on modern embedded systems. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2013. LNCS, vol. 8112, pp. 491–498. Springer, Heidelberg (2013)
6. Drusinsky, D., Harel, D.: Using statecharts for hardware description and synthesis. IEEE Trans. Computer-Aided Design 8(7), 798–807 (1989)
7. Gomes, L., Costa, A.: From use cases to system implementation: Statechart based co-design. In: 1st ACM & IEEE Conf. Formal Methods and Programming Models for Codesign, MEMOCODE 2003, Mt St-Michel, France, pp. 24–33 (2003)
8. Doligalski, M.: UML state machine implementation in FPGA devices by means of dual model and Verilog. In: 11th IEEE Inter. Conf. Industrial Informatics, INDIN 2013, Bochum, Germany, pp. 177–184 (2013)
9. Murata, T.: Petri nets: Properties, analysis and applications. Proc. IEEE 77(4), 541–580 (1989)
10. Karatkevich, A.: Dynamic Analysis of Petri Net-Based Discrete Systems. LNCIS, vol. 356. Springer, Berlin (2007)
11. Biliński, K., Adamski, M., Saul, J., Dagless, E.: Petri-net-based algorithms for parallel-controller synthesis. IEE Proc. Comput. Digit. Tech. 141(6), 405–412 (1994)
12. Esparza, J., Silva, M.: On the analysis and synthesis of free choice systems. In: Rozenberg, G. (ed.) Advances in Petri Nets 1990. LNCS, vol. 483, pp. 243–286. Springer, Heidelberg (1991)
13. Barkaoui, K., Pradat-Peyre, J.F.: On liveness and controlled siphons in Petri nets. In: Billington, J., Reisig, W. (eds.) ICATPN 1996. LNCS, vol. 1091, pp. 57–72. Springer, Heidelberg (1996)
14. Cortés, L.A., Eles, P., Peng, Z.: Modeling and formal verification of embedded systems based on a Petri net representation. J. Sys. Arch. 49(12-15), 571–598 (2003)
15. Girault, C., Valk, R.: Petri Nets for System Engineering: A Guide to Modeling, Verification, and Applications. Springer, Berlin (2003)
16. Węgrzyn, M.: Implementation of safety critical logic controller by means of FPGA. Ann. Rev. Control 27(1), 55–61 (2003)
17. Bukowiec, A., Węgrzyn, M.: Design of Safety Critical Logic Controller Using Devices Integrated Microprocessor with FPGA. In: Proc. SPIE, vol. 5775, pp. 377–384 (2005)
18. Tkacz, J., Adamski, M.: Macrostate encoding of reconfigurable digital controllers from topological Petri net structure. Przegląd Elektrotechniczny 88(8), 137–140 (2012)
19. Bukowiec, A., Adamski, M.: Synthesis of Petri nets into FPGA with operation flexible memories. In: IEEE 15th Int. Symposium Design and Diagnostics of Electronic Circuits and Systems, DDECS 2012, Tallinn, Estonia, pp. 16–21 (2012)
20. Gniewek, L., Kluska, J.: Hardware implementation of fuzzy Petri net as a controller. IEEE Trans. Syst., Man, Cybern. B 34(3), 1315–1324 (2004)
21. Adamski, M., Węgrzyn, M.: Petri nets mapping into reconfigurable logic controllers. Electronics and Telecommunications Quarterly 55(2), 157–182 (2009)
22. Karatkevich, A., Wisniewski, R.: Wyznaczanie pokrycia sieci Petriego przez SM-komponenty z wykorzystaniem przeszukiwania grafów. Przegląd Elektrotechniczny 88(8), 141–144 (2012)
23. Savi, V., Xie, X.: Liveness and boundedness analysis for Petri nets with event graph modules. In: Jensen, K. (ed.) ICATPN 1992. LNCS, vol. 616, pp. 328–347. Springer, Heidelberg (1992)
24. Sałamaj, M.: Conception of a control unit for critical systems. Intl. Journal of Electronics and Telecommunications 59(4), 363–368 (2013)
25. Halang, W.: Automated control systems for the safety integrity levels 3 and 4. Comput. Syst. Sci. Eng. 24(1) (2009)

A Method of Determining an Electric Energy Meter Maximum Uncertainty

Piotr Mróz, Andrzej Olencki, and Arkadiusz Bukowiec

University of Zielona Góra, Institute of Computer Engineering and Electronics,
ul. Licealna 9, 65-417 Zielona Góra, Poland
{p.mroz,a.olencki,a.bukowiec}@iie.uz.zgora.pl

Abstract. This paper proposes the adoption of the estimator of the standard deviation as a quality measurement of the test. The uncertainty type A can anticipate and adopt its value a priori. Imposition of any noise during the test should be within the revalue value of s_{max}. Knowing the characteristics of s_{max} as a function of time, it is possible to determine the minimum duration of the test while maintaining the required uncertainty.

Keywords: energy meter testing, accuracy of energy meter, repeatability of measurements, standard deviation, measurement uncertainty.

1 Introduction

For many years, there has been, in principle, the compatibility of views on this, in which measuring points (also called points of a load meter) energy meters, both inductive and electronic (static) [1], should be checked. The standards and regulations [2-4] specified the tables describing the maximum permissible errors counters for individual measurement points during type approval and verification of the meter.

It is very important to provide the required reproducibility of the test during testing both using energy meter test system [5], as well as portable testers [6]. The best indicator of the repeatability is uncertainty. An exemplary scheme of the testing system is shown in Figure 1.

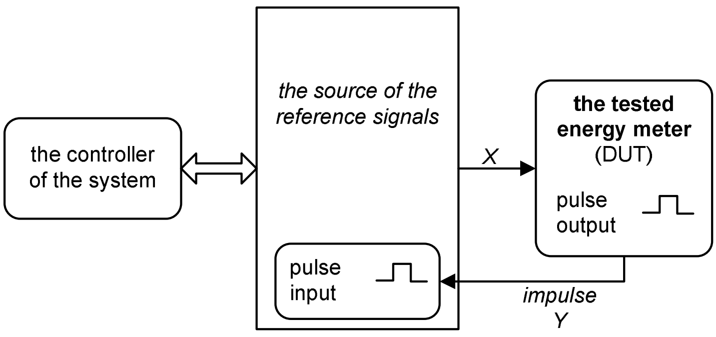


Fig. 1. An exemplary scheme of the testing system

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_57, © Springer-Verlag Berlin Heidelberg 2014

Typically it consists of:

- the controller of the system,
- a source of the reference signals,
- the tested energy meter (DTU).

The controller of the system is a PC computer with dedicated software. It could be also implemented as dedicated hardware controller [7-9]. There can be considered several models for control algorithm, like finite state machines [10, 11], Petri nets [12, 13], or UML state machines [14, 15]. The source of the reference signals is a three phase power calibrator. There is used Calmet C300 three and single phase power and energy calibrator [16] in our testing system. the tested energy meter (DTU) is a device under test. Conducted researches have been worked out with use of LUMEL EM03 energy meter [17].

2 Determination of Uncertainty

The expanded percentage uncertainty U for a directly measured error of the energy meter (DUT), when the uncertainties of type A and type B are taken into consideration, results from:

$$U = k \cdot u_C , \tag{1}$$

where k is the coverage factor and uC is the combined standard uncertainty:

$$u_C = k \cdot \sqrt{u_A^2 + u_B^2} , \tag{2}$$

The standard uncertainty of type A – uA is calculated from the observed scatter of the DUT's errors in the measurements series for each test point and may be expressed as the estimation of standard deviation:

$$u_A = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(E_i - \overline{E}\right)^2} , \tag{3}$$

where n is a number of measurements for each test point, Ei is the error percentage [3] in relation to the measurement value for the ith measurement point and $\overline{E}$ is the average value of the percentage error. The standard uncertainty of type A should be particularly taken into account if there is a need for efficient energy meters verification, which requires energy measurement time shortening – in effect, it increases the standard deviation value.

The reasons for the uniqueness of the energy meter error measurement results lie both in the meter and in the MTE tester. According to the standard [18], the tester should have about ten times better reproducibility than the tested meter. Therefore, when using the relevant tester, reproducibility of the measurement results depends on the properties of the counter, and in particular on the so-called inequality of the pulse output of the meter.

3 Determination of the Maximal Value of Standard Deviation

The inequality of the pulse output is not specified in the technical documentation of the meter. The non-uniformity is manifested in such a way that, even with the provision the load counter constancy in function of time, the output pulses occur at irregular intervals. The typical case is as follows: some pulses appear at regular intervals, then another pulse appears at some other interval – more or less, followed by a further cycle at regular intervals. Of course, we are talking about electronic counters, because in the case of electromechanical electricity meters, pulse output inequality results from the inequality of meters shield rotation and delays of an electromechanical system that emits electrical impulses.

The inequality of the pulse output of the meter becomes important when shortening the time of energy measurement in order to reduce the time of meter testing. During short time of energy measurement it is possible that the energy measurements are performed only during the appearance of pulses at regular intervals, and then the standard deviation of the values εS approaches zero, as shown in Table 1. For high values of the load of the meter EM03 (symmetric load 50 A at cos $\varphi = 1$), the measurement of the counter error was performed ten times in the automatic test with a very short measurement time tM = 1 s and each with three repetitions. The estimate of the standard deviation s (marked in the Table 1 as εS) received successively the values: 0.000%, 0.000%, 0.000%, 0.055%, 0.000%, 0.000%, 0.000%, 0.055%, and so on.

Appearing from time to time larger values of the standard deviation s (in Table 1 marked as εs). The maximal value of standard deviation s were indicated as smax. The relationship of smax as a function of time measurement is shown in Figure 2. It is determined individually for each tested electricity meter. The horizontal line shows the maximum value of required standard deviation for device testing. The figure shows that to meet the requirement of [18] for the permissible value of the estimate of the standard deviation, the measurement time of energy tM should not be less than three seconds. This result demonstrates the high quality of the tested meter in the sense of a small inequality of its pulse output which allows using the time efficient procedures of testing.

Table 1. The example measurements results

I2 [A]	I3 [A]	F [Hz]	Phi1	Phi2	Phi3			Limit [%]	ε [%]	ε_s [%]	OK
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.368	0.055	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.368	0.055	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓
50.000	50.000	50.000	Cos 1.00 L	Cos 1.00 L	Cos 1.00 L		L123	1.000	-0.336	0.000	✓

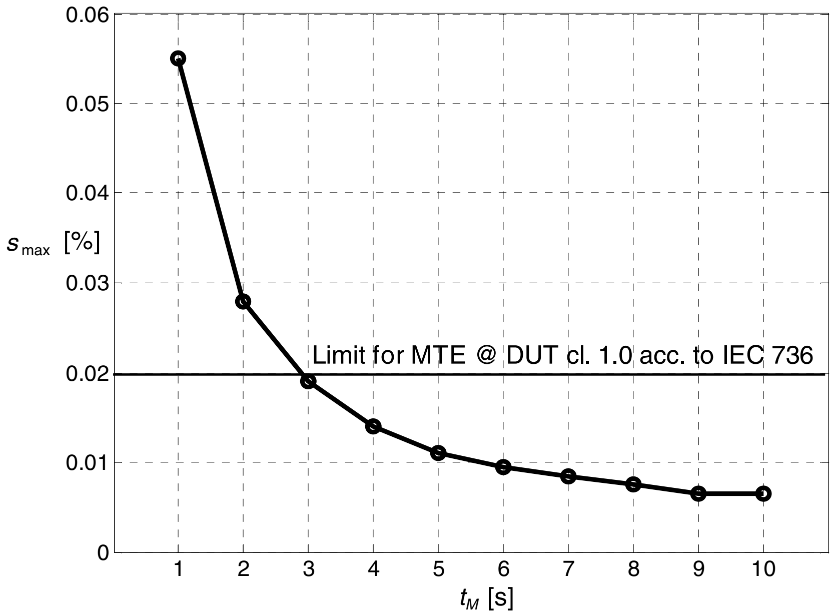


Fig. 2. Diagram of the maximum value of the standard deviation estimator as a function of time measurement

In order to specify the value of the uncertainty type A we suggest to assume that

$$u_A = s_{\max} \,. \tag{4}$$

So accepted value will be overvalued in relation to the actual value of uA. In consequence it will extend the testing time insignificantly, but it will ensure the compliance with the established criteria of the test.

4 Conclusions

This paper proposes the adoption of the estimator of the standard deviation as a quality measurement of the test. The uncertainty type A can be anticipated and its value can be adopted a priori. Imposition of any noise during the test should be within the revalued value of smax. Knowing the characteristics of smax as a function of time, it is possible to determine the minimum duration of the test while maintaining the required uncertainty.

References

1. Mróz, P., Urbański, K., Szmytkiewicz, J.: Principles of testing meters and power quality analyzers. Measurement Automation and Monitoring 55(7), 415–417 (2009)
2. President of Polish Committee for Standardization and Quality Measures: About tools verification to measuring electricity. Instruction No. 5, Polish Committee for Standardization and Quality Measures, Warsaw (1981)
3. President of Central Office of Measures: Instruction of checking useful and reference of energy meters AC. Instruction No. 5, Central Ofice of Measures, Warsaw (1995)
4. Minister of Economy, Labour and Social Policy: Metrological requirements to be met by active electrical energy meters alternating current accuracy class 0.2, 0.5, 1, and 2. Regulation, Ministry of Economy, Labour and Social Policy, Warsaw (2004)
5. Olencki, A., Urbański, K., Szmytkiewicz, J.: Stationary single position meter test stations. Przegląd Elektrotechniczny 85(7), 189–191 (2009)
6. Olencki, A.: Development of power network parameters standards. Measurement Automation and Monitoring 55(12), 993–996 (2009)
7. Milik, A., Hrynkiewicz, E.: Reconfigurable logic controller, architecture, programming, implementation. In: Ciazynski, W., Hrynkiewicz, E., Klosowski, P. (eds.) A Proceedings Volume from the 5th IFAC Workshop on Programmable Devices and Systems, PDS 2001, pp. 163–168, Pergamon, London (2002)
8. Borowik, G., Rawski, M., Łabiak, G., Bukowiec, A., Selvaraj, H.: Efficient logic controller design. In: Fifth International Conference on Broadband and Biomedical Communications, IB2Com 2010, Malaga, Spain (2010) [CD-ROM]
9. Adamski, M., Wiśniewska, M., Wiśniewski, R., Stefanowicz, L.: Application of hypergraphs to the reduction of the memory size in the microprogrammed controllers with address converter. Przegląd Elektrotechniczny 88(8), 134–136 (2012)
10. Barkalov, A., Titarenko, L.: Logic Synthesis for FSM-based Control Units. LNEE, vol. 53. Springer, Berlin (2009)
11. Bukowiec, A.: Synthesis of FSMs based on architectural decomposition with joined multiple encoding. International Journal of Electronics and Telecommunications 58(1), 35–41 (2012)
12. Tkacz, J., Adamski, M.: Macrostate encoding of reconfigurable digital controllers from topological Petri net structure. Przegląd Elektrotechniczny 88(8), 137–140 (2012)

13. Bukowiec, A., Adamski, M.: Synthesis of Petri nets into FPGA with operation flexible memories. In: Proceedings of the IEEE 15th International Symposium on Design and Diagnostics of Electronic Circuits and Systems, DDECS 2012, Tallinn, Estonia, pp. 16–21 (2012)
14. Gomes, L., Costa, A.: From use cases to system implementation: Statechart based co-design. In: Proceedings of 1st ACM & IEEE Conference on Formal Methods and Programming Models for Codesign, MEMOCODE 2003, Mont Saint-Michel, France, pp. 24–33. IEEE Computer Society Press (2003)
15. Doligalski, M., Adamski, M.: UML state machine implementation in FPGA devices by means of dual model and Verilog. In: 11th IEEE International Conference on Industrial Informatics, INDIN 2013, Bochum, Germany, pp. 177–184. IEEE (2013)
16. Calmet: C300 three phase power calibrator and power engineering devices tester. Data Sheet EN 2013-03, Calmet, Zielona Góra (2013)
17. Lumel: EM03 Electronic Watt-hour Meter, Zielona Góra (2012)
18. CEI IEC 736: Testing equipment for electrical energy meters. Technical Report IEC/TR 60736 ed1.0, International Electrotechnical Commission, Genève (1982)

Self-reconfigurable Logic Controller Architecture

Michał Doligalski

University of Zielona Góra, Institute of Computer Engineering and Electronics,
ul. Licealna 9, 65-417 Zielona Góra, Poland
m.doligalski@iie.uz.zgora.pl

Abstract. Logic controller design and implementation by means of FPGA devices is a quite common solution including partial reconfiguration functionality. The possibility of the PR is provided by the decomposition of the logic controllers into functional logic blocks. The ability of reconfiguration of the dynamic blocks increases the functionality of the logic controller and can reduce required hardware resources.

The self-reconfiguration logic controllers consists of two blocks. Static master control unit (MCU) and partial logic controller (PLC). The MCU is responsible for partial reconfiguration control. The PLC block is responsible for the realisation of control algorithm. This solution is especially useful in the case of complex systems, for which variants of the control program are implemented with the use of the same set of resources (sensors and actuators).

The paper presents architecture of the self-reconfigurable logic controller.

Keywords: RLC, self-reconfiguration, partial reconfiguration, UML, petri Net, FPGA, state machine.

1 Introduction

A logic controller (LC) (Fig. 1 a) can be perceived as devices controlling industrial processes or as the basic element of integrated information microsystems in the form of a special-purpose digital systems [3, 12, 16].

The communication with a controlled object is provided through input of X and output of Y signals. Communication with a system operator (or other external systems) is performed through external X_E input signals and external Y_E output signals [16]. Logic controllers can also be an element of integrated digital systems in which a data path (functional blocks) and a control path (a control unit) are distinguished [2]. Decomposition of the system into a Control unit (CU) and an operational unit (OU) [2] not only increases the comfort of development but also facilitates a reuse of data processing modules that were developed earlier or IP core. The paper will consider LCs dedicated to industrial applications.

Petri nets and UML state machines are popular method of the specification, and the synthesis can be done automatically into HDL (Verilog and VHDL languages) [6]. Alternative methods of the specification by means of dual model [7, 8] simplify the design process. Thanks to formal analysis methods, LC specified by dual model can be verified towards design faults detection [11].

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_58, © Springer-Verlag Berlin Heidelberg 2014

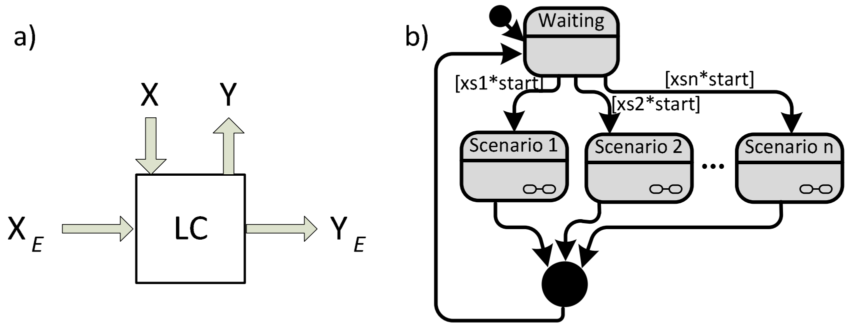


Fig. 1. Logic controller block (a), UML state machine for alternative LC scenarios (b)

The implementation of the logic controller with the use of Field Programmable Gate Array devices (FPGA) [5, 16, 17] is one of the possible solution , beside other ways e.g. the ones with use of embedded systems or microprocessors [1].

Reconfigurable devices (FPGA) can be configured by bitstreams stored in external memory. Typically during reconfiguration the process covers the whole area of FPGA device. However, there is also a possibility for partial reconfiguration where some parts remains unchanged and some parts are reconfigured.

Logic controllers witch alternative scenarios (ASLC) are the special case of logic controller. Alternative scenario occurs when for particular control process and for the same set of inputs (X, X_E) and the same set of outputs (Y, Y_E) the behavior of the control process may be performed by different ways. Such situation was presented in Fig. 1 b. At the top level of the logic controller, composite states corresponds to alternative scenario. The scenario is executed by signals: *xs1*, *xs2*, ..., *xsn*. There is no concurrency at the top level description. Topically signals responsible for scenario selection may be included into the set of X_E signal.

Each scenario is implemented at the second level (and bellow). Direct implementation of such hierarchical LC requires the reservation of the FPGA device resource for each scenario even though, at the same time only one scenario can be executed.

The paper presents decomposition of the ASLC and the application of partial reconfiguration for the alternative scenarios implementation.

2 Partial Reconfiguration

In classical approach, full static reconfiguration requires reload of whole FPGA device configuration which causes control interruption. The scenario of the control algorithm cannot be switched without direct implementation. It results in the size growth of required resources. There is no possibility to add other scenarios after bitstream generation, but it is very possible that at the specification stage, not all scenarios will be known and further update will be required. Also alternative scenario may be introduced after the control system deployment and treated as a system improvement. So there will be no possibility of the implementation of all alternative scenarios in one control algorithm. The other issue of direct scenario implementation results from the size of the algorithm. Each context will increase space required for the implementation.

Partial reconfiguration (PR) is the feature of modern FPGA devices that provides control switching. Some parts of the devices may be reconfigured, while others will be preserved. The paper will consider dynamic reconfiguration when changes in the configuration are provided in run-time. There are two approaches for module- based partial reconfiguration (MBPR) and difference based partial reconfiguration (DBPR).

In the DBPR oriented developing process, two configurations are compared by external tool provided by FPGA vendor and, as a result, differential configuration file is generated. This approach is dedicated rather to small design changes e.g. Block Ram contents modification [15].

Full benefits from the partial-oriented design comes with modular based partial reconfiguration (MBPR). The static (SP) and reconfigurable partitions (RP) are identified within the design. For each RP, set of alternative reconfigurable modules (RMs) is developed. In the proposed solution each scenario will be implemented as one RMs. The bottom-up synthesis is performed and each module is synthesized separately. Finally top-level module connect both static and reconfigurable modules - SRLC module. During dynamic module-based partial reconfiguration, structure of separated reconfigurable partition may be changed without affecting the operation of other modules. The paper presents revised approach for PR design of logic controllers. Previous methodology increases design complexity [10]. It requires the application of *bus macro* between reconfigurable module and other modules. Bus macro was an interface between dynamic modules and other modules (both static and dynamic), implemented by means of three-state buffers. In the new modern FPGA devices and updated CAD software (Plan Ahead) it's not required to add bus-macro inside the design. The management of the partitions and its configurations (reconfigurable modules) is transparent and intuitive. It is necessary that each RM will have the same interface and space reserved for reconfigurable partition will be sufficient to implement the biggest reconfigurable module. The configuration files can be stored in external flash memory and be accessed by microprocessor or internal module responsible for partial reconfiguration.

3 Self-reconfigurable Architecture

To optimize logic resources, ASLC was decomposed into two functional blocks: an alternative logic controller (ALC) and a master logic controller (MLC) (Fig. 2). ALC is responsible for scenario implementation control, MLC is responsible for scenario switching. For this architecture ALC block are treated as a reconfigurable portion and the MLC block as a static module. For the initial configuration the ALC block can be implemented as a no-op (bank module). The optimization is achieved by a separate implementation of the scenario as reconfigurable modules. Each reconfigurable module for partition ALC, represents one scenario. The ALC block responsible for control process implementation communicates with the MLC block by signals E, R, F. The E signal is responsible for initiation of the control, R is for reset purpose and F communicates so that the scenario can be finished.

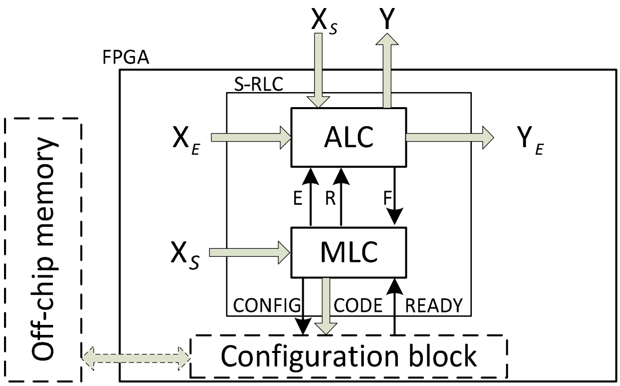


Fig. 2. Self-reconfigurable logic controller

The behavior of functional blocks, both ALC and MLC was described by means of UML state machine diagram (Fig. 3). Each scenario is specified as a separate composite state (Fig 3 b). The state is entered when transition guarded by E condition is fired. The completion of the state machine is indicated by F signal. In the presented approach it is not possible to switch context when scenario is executed, however reset signal (R) can be treated as the exception trigger and scenario interruption can be provided. Exceptions handling mechanism and its specification by means of UML state machine was presented in [9].

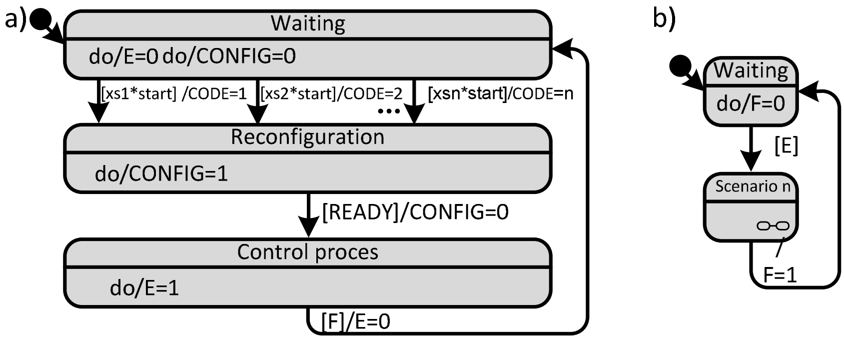


Fig. 3. UML state machine for MLC (a), UML state machine for scenario (b)

The behavior of the MLC is specified by a three-state UML state machine. The MLC block is responsible for decoding proper configuration (*CODE* signal) and initiation of the partial reconfiguration (*CONFIG* signal). When the ALC unit is configured properly, the MLC initiates state machine that is responsible for an alternative control process execution. After completion of the control algorithm, state of the MLC unit changes into a waiting mode for new execution. For the proposed method each scenario execution will be preceded by reprogramming of the ALC unit. It is possible to avoid reprogramming when in respect of subsequent startups the same scenario is selected.

The MLC unit communicates with configuration block by means of *CONFIG* and *CODE* signals as well as *READY*. *CONFIG* signal triggers partial reconfiguration of the ALC unit. Reconfiguration is performed by the configuration block. Proper configuration (*CODE*) is chosen by MLC unit and may be interpreted as offset in off-chip memory, in this case binary code is most obvious. The proper partial reconfiguration of the FPGA devices will results in ALC modification and will be indicated by *READY* signal.

The configuration block shall also implement exceptions handling mechanism. In particular, it should include bitstream CRC checking or reconfiguration faults detection. The exceptions shall be handled and, if necessary, MLC shall be notified. The implementation of the configuration block including exceptions handling mechanism is beyond the scope of this article. The proposed architecture is platform-independent, the variety of implementation methods were described in other papers [4, 13]. The idea is to use off-chip memory to reduce internal resources. In the case of bidirectional memory, scenarios can be updated without the interruption of control process execution. The memory content can be swapped with a new scenario - but only when the interface of the ALC remains unchanged. It's essentially useful in the case of LC improvements or faults removing.

4 Summary

In the paper proposition of the new architecture of the logic controller was presented. Both units (MLC and ALC) were described by means of UML state machine diagrams. Proposed architecture is dedicated to logic controllers with regard to which different scenario exists and are chosen at the beginning of the control process. The scenarios can be implemented even after implementation and system deployment and inserted into off-chip memory. This functionality is limited by the ALC interface and the reconfigurable partition size. There is no possibility to provide direct changes in the interface without re-implementation of the whole system. The resizing of the reconfigurable partition is not possible either. The logic controller is decomposed and one module is assigned to reconfigurable partition. This module is responsible for scenario implementation and will be partially reconfigured.

The biggest advantage of the proposed method is the independence from FPGA devices families. It was achieved by the decomposition into two blocks, whereby one of the blocks communicates with external configuration blocks by simple interface. Further research will consider faults detection and automatic ASLC reconfiguration upon errors.

References

1. Adamski, M., Wiśniewska, M., Wiśniewski, R., Stefanowicz, L.: Application of hypergraphs to the reduction of the memory size in the microprogrammed controllers with address converter. Przegląd Elektrotechniczny 88(8), 134–136 (2012)

2. Baranov, S.I.: Logic Synthesis for Control Automata. Kluwer Academic Publishers, Norwell (1994)
3. Barkalov, A., Titarenko, L.: Logic Synthesis for Compositional Microprogram Control Units. LNEE, vol. 22. Springer, Heidelberg (2008)
4. Bayar, S., Yurdakul, A.: Self reconfiguration on Spartan-III FPGAs with compressed partial bitstreams via a parallel configuration access port (cPCAP) core. Research in Microelectronics and Electronics 2008, 137–140 (2008)
5. Bukowiec, A., Doligalski, M.: Petri net dynamic partial reconfiguration in FPGA. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST. LNCS, vol. 8111, pp. 436–443. Springer, Heidelberg (2013)
6. Bukowiec, A., Mróz, P.: An FPGA synthesis of the distributed control systems designed with Petri nets. In: Proc. IEEE 3rd International Conference on Networked Embedded Systems for Every Application, NESEA 2012, Liverpool, United Kingdom, vol. 6 (2012)
7. Doligalski, M., Adamski, M.: UML state machine implementation in FPGA devices by means of dual model and Verilog. In: 11th IEEE International Conference on Industrial Informatics, INDIN 2013, Bochum, Germany, pp. 177–184 (2013)
8. Doligalski, M.: UML state machine conversion in field of dual specification (in Polish). Przegląd Elektrotechniczny 85(7), 192–195 (2009)
9. Doligalski, M., Adamski, M.: Exceptions and deep history state handling using dual specification (in Polish). Przegląd Elektrotechniczny 86(9), 123–125 (2010)
10. Doligalski, M., Węgrzyn, M.: Partial reconfiguration-oriented design of logic controllers. In: SPIE: Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2007, vol. 6937, p. 10 (2007)
11. Gaczyński, A., Tkacz, J., Doligalski, M.: Concept and design of Gentzen deduction system in RDBMS - IBM DB2. Przegląd Elektrotechniczny 88(8), 148–151 (2012)
12. Hrynkiewicz, E., Milik, A., Mocha, J.: Dynamically reconfigurable concurrent implementation of the binary control (in Polish). Electronics: Constructions, Technologies, Applications 49(11), 187–190 (2008)
13. Lai, V., Diessel, O.: ICAP-I: A reusable interface for the internal reconfiguration of Xilinx FPGAs. In: International Conference on Field-Programmable Technology 2009 – FPT 2009, pp. 357–360 (2009)
14. Rawski, M., Tomaszewicz, P., Borowik, G., Łuba, T.: 5 logic synthesis method of digital circuits designed for implementation with embedded memory blocks of fPGAs. In: Adamski, M., Barkalov, A., Węgrzyn, M. (eds.) Design of Digital Systems and Devices. LNEE, vol. 79, pp. 121–144. Springer, Heidelberg (2011)
15. Rawski, M., Borowik, G., Łuba, T., Tomaszewicz, P., Falkowski, B.: Logic synthesis strategy for FPGAs with embedded memory blocks. In: 16th International Conf. Mixed Design of Integrated Circuits Systems 2009 MIXDES, pp. 296–301 (2009)
16. Tkacz, J., Adamski, M.: Logic design of structured configurable. In: Proc. IEEE 3rd International Conference on Networked Embedded Systems for Every Application NESEA 2012, Liverpool, United Kingdom, p. 6 (2012)
17. Doligalski, M., Bukowiec, A.: Partial Reconfiguration in the Field of Logic Controllers Design. Intl Journal of Electronics and Telecommunications 59(4), 351–356 (2013)

Data Synchronization Operations for a Monitoring System with a Very Large Number of PHDs

KeeHyun Park, SeungHyeon Lim, and Joonsuu Park

Department of Computer Engineering, Keimyung University, Daegu, Republic of Korea
{khp,burningwing,suuyapark}@kmu.ac.kr

Abstract. Various PHDs (Personal Healthcare Devices) have been developed to enable the easy monitoring of individual bio-signals. But while most of the related studies have dealt only with 1:1 communication between a PHD and a monitoring server, serious bottleneck problems at the monitoring server are inevitable in real world in which a very large number of PHD users are monitored. In this paper, data synchronization operations for a hierarchical monitoring system to be applied when a very large number of PHD clients are used are proposed. Detailed execution mechanisms for data synchronization operations are described. After every component uses XOR operations to calculate the difference between what it had in the past and what it has now, it stores the difference in its own table and sends the difference to the component in the upper layer in order to reduce storage/communication costs.

Keywords: Personal healthcare device, Data synchronization, Hierarchical system, XOR operations.

1 Introduction

In recent years, ubiquitous technologies have been developed to enable the remote monitoring of personal health, from any location and at any time. Therefore, various PHDs (Personal Healthcare Devices) have been developed to monitor individual bio-signals [1-5]. However, since most of the related studies have considered only 1:1 communication between a PHD and a monitoring server [6-10], they have not addressed the serious bottleneck problems at the monitoring server that are inevitable in real world applications, in which a very large number of PHD users are being monitored.

A PHD data synchronization system can be defined as a system that consists of some PHDs and a monitoring server. A gateway can be located between PHDs and a monitoring server when more than one communication protocol is being used [10]. The ISO/IEEE 11073 standard [1-5] was proposed to define how personal health data should be exchanged between a PHD and a monitoring server. When remote PHD management is needed, data synchronization/device management protocols can be used along with the ISO/IEEE 11073 communication protocol [6-10]. However, most of the studies mentioned above deal with a 1:1 communication mechanism between a

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_59, © Springer-Verlag Berlin Heidelberg 2014

PHD and a monitoring server, which fails to address the serious bottleneck problems that can occur at the monitoring server when a very large number of PHD users have to be monitored.

In this paper, data synchronization operations in a hierarchical system in which a very large number of PHD clients are used are proposed. For better workload distribution, a hierarchical data synchronization system is proposed for a very large number of PHD clients. To reduce storage/communication overhead, XOR-based data synchronization mechanisms between components in the system are proposed. Every component of the system uses XOR operations to calculate the difference between what it had in the past and what it has now. It stores the difference and sends the difference to the component in the upper layer in order to reduce storage/communication costs. The remainder of this paper is organized as follows. Section 2 proposes data synchronization operations for a hierarchical monitoring system for a very large number of PHD clients. Section 3 draws some conclusions and proposes some directions for future research.

2 Data Synchronization Operations with a Very Large Number of PHDs

2.1 System Overview

Figure 1 shows the structure of a data synchronization system with a very large number of diverse PHDs. The system consists of Mobile devices (Ms), Group servers (Gs), Regional servers (Rs) and a Global server (GB). The Mobile devices represent PHDs which capture users' biomedical signals or medication status. PHDs are grouped together to form a group.

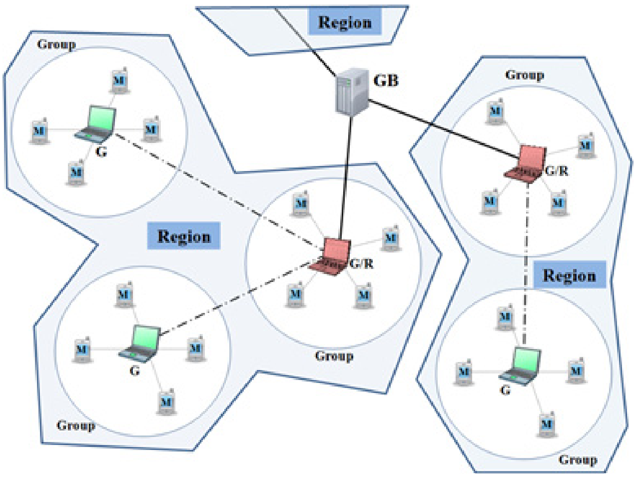


Fig. 1. The structure of a data synchronization system with a very large number of PHDs

Once biomedical signals or medication status are captured by a Mobile device (PHD), they are sent to the Group server to which the mobile device belongs. The Group server plays the role of the data synchronization server in the group. Data synchronization operations are performed between the Group server and the Mobile devices. Group servers are grouped together to form a region. Group servers process data transmitted from Mobile devices and send it to the Regional server to which they belong. The Regional server plays the role of the data synchronization server in the region and device management operations are performed between the Regional server and the Group servers. One of the Group servers in the same region can be the Regional server (G/R) in the region. There is one and only one Global server in the integrated data synchronization system. Again, Regional servers process data transmitted from the Group servers and send it to the Global server. The Global server acts as the uppermost data synchronization server in the system and data synchronization operations are performed between the Global server and the Regional servers.

2.2 Definitions and Notations

Throughout this paper, the following definitions and notations are used:

- Mi,j : i-th Mobile device in j-th group which has l Mobile devices.
- Gj,k : j-th Group server in k-th region which has m Group servers.
- Rk : k-th Regional server in the system which has n Regional servers.
- GB : the Global server in the system.
- LocalTable : A data table of entries (LocalUID, Data) stored in a Mobile device whose unique ID is LocalUID .
- GroupTable : A data table of entries (DeviceID, LocalUID, Data) stored in a Group server whose unique ID is (DeviceID, LocalUID).
- XOR_GroupTable : An XOR-ed data table of entries (GroupUID, XOR-ed Data) in the Group server whose unique ID is GroupUID. This table is obtained by performing XOR operations on all the data in a GroupTable in terms of LocalUID. This table is transmitted to the related Regional server.
- XOR_RegionTable : An XOR-ed data table of entries (RegionUID, XOR-ed Data) in the Regional server whose unique ID is RegionUID. This table is obtained by performing XOR operations on all the data in XOR_GroupTables, which are received from Group servers, in terms of GroupUID. This table is transmitted to the Global server.
- XOR_GlobalTable : An XOR-ed data table of entries (GlobalUID, XOR-ed Data) in the Global server whose unique ID is GlobalUID. This table is obtained by performing XOR operations on all the data in XOR_RegionTables, which are received from Regional servers, in terms of RegionUID.
- LData_p_Mi,j(t) : 2-tuple data (LocalUID, data) in the LocalTable of Mi,j whose LocalUID is p at time t.
- Packet_LData_p_Mi,j(t) : Communication packet containing LData_p_Mi,j(t) transmitted from Mi,j to Gj,k at time t.
- GData_i_p_G,j,k(t) : 3-tuple data (DeviceID, LocalUID, data) in the GroupTable of Gj,k whose DeviceID is i and LocalUID is p at time t.
- XGData_p_Gj,k(t) : 2-tuple data (GroupUID, XOR-ed data) in the XOR_GroupTable of Gj,k whose GroupUID is p at time t.
- Packet_XGData_p_Gj,k(t) : Communication packet containing XGData_p_Gj,k(t) transmitted from Gj,k to Rk at time t.
- XRData_p_Rk(t) : 2-tuple data (RegionUID, XOR-ed data) in the XOR_RegionTable of Rk whose RegionUID is p at time t.
- Packet_XRDatap_Rk(t) : Communication packet containing XRData_p_R,k(t) transmitted from Rk to the Global server at time t.

- XGBData_p_GB(t) : 2-tuple data (GlobalUID, XOR-ed data) in the XOR_GlobalTable of the Global server whose GlobalUID is p at time t.
- Ij(t) : a set of Mi,js whose data is updated at time t. (i = 1, 2, …, l)
- Jk(t) : a set of Gj,ks whose data is updated at time t. (j = 1, 2, …, m)
- K(t) : a set of Rks whose data is updated at time t. (j = 1, 2, …, n)

2.3 XOR-Based Data Synchronization Operations

ADD Operation

– At Mobile device Mi,j (for i $\in$ Ij(t)): Upon capturing new signals, a Mobile device generates its own data from the signals and sends the data to the related Group server.

- Generates LData_p_Mi,j(t) and stores it in its LocalTable, for every p
- Generates Packet_LData_p_Mi,j(t) and sends it to Gj,k

– At Group server Gj,k: Upon receiving data from a Mobile device, a Group server generates its own data to store it in its GroupTable. Also, the Group server calculates the differences between the data received from the Mobile device and the data stored previously in its XOR_GroupTable in order to store the differences in its XOR_GroupTable. In addition, the Group server calculates the differences between new data and previous data in its XOR_GroupTable in order to send the differences to the related Regional server.

- Receives Packet_LData_p_Mi,j(t) and generates GData_i_p_Gj,k(t), for every p
- Stores GData_i_p_Gj,k(t) in its GroupTable
- Generates XGData_p_Gj,k(t) (= XGData_p_Gj,k(t-1) $\oplus$ Packet_LData_p_Mi,j(t)) and store it in its XOR_GroupTable
- Generates Packet_GData_p_Gj,k(t) (= XGData_p_Gj,k(t) $\oplus$ XGData_p_Gj,k(t-1)) and send it to Rk

– At Regional server Rk: One of Group servers is designated as a Regional server. Upon receiving data from a Group server, a Regional server calculates the differences between the data received from the Group server and the data stored previously in its XOR_RegionTable in order to store the differences in its XOR_RegionTable. In addition, the Regional server calculates the differences between new data and previous data in its XOR_RegionTable in order to send the differences to the Global server.

- Generates XRData_p_Rk(t) (= XRData_p_Rk(t-1) $\oplus$ Packet_GData_p_Gj,k(t) (for j $\in$ Jk(t)) and store it in its XOR_RegionTable
- Generates Packet_XRData_p_Rk(t) (= Datap_Rk(t) $\oplus$ Datap_Rk(t-1)) and send it to the Global server

– At the Global server: Upon receiving data from a Regional server, the Group server calculates the differences between the data received from the Regional server and the data stored previously in its XOR_GlobalTable

- Generates XGBData_p_GB(t) (= XGBData_p_GB(t-1) $\oplus$ Packet_XRData_p_Rk(t) (for k $\in$ K(t)) and store it in its XOR_GlobalTable

Replace Operation

- At Mobile device Mi,j (for i $\in$ Ij(t)): Upon acknowledging an update of a data, a Mobile device applies the update to the specific data and sends the data to the related Group server.

- Generates LData_p_Mi,j(t) and restores it in its LocalTable, for every p
- Generates Packet_LData_p_Mi,j(t) and sends it to Gj,k

- At Group server Gj,k: Upon receiving an updated data from a Mobile device, a Group server restores the related data stored in its GroupTable. Also, the Group server calculates the differences between the data received from the Mobile device and the data stored previously in its XOR_GroupTable in order to store the differences in its XOR_GroupTable. In addition, the Group server calculates the differences between new data and previous data in its XOR_GroupTable in order to send the differences to the related Regional server.

- Receives Packet_LData_p_Mi,j(t) and generates GData_i_p_Gj,k(t), for every p
- Restores GData_i_p_Gj,k(t) in its GroupTable
- Generates XGData_p_Gj,k(t) (= GData_i_p_Gj,k(t-1) $\oplus$ Packet_LData_p_Mi,j(t)) and restore it in its XOR_GroupTable
- Generates Packet_GData_p_Gj,k(t) (= XGData_p_Gj,k(t) $\oplus$ XGData_p_Gj,k(t-1)) and send it to Rk

- At Regional server Rk; Upon receiving data from a Group server, a Regional server calculates the differences between the data received from the Group server and the data stored previously in its XOR_RegionTable in order to store the differences in its XOR_RegionTable. In addition, the Regional server calculates the differences between new data and previous data in its XOR_RegionTable in order to send the differences to the Global server.

- Generates XRData_p_Rk(t) (= XRData_p_Rk(t-1) $\oplus$ Packet_GData_p_Gj,k(t) (for j $\in$ Jk(t)) and restore it in its XOR_RegionTable
- Generates Packet_XRData_p_Rk(t) (= XRData_p_Rk(t) $\oplus$ XRData_p_Rk(t-1)) and send it to the Global server

- At the Global server: Upon receiving data from a Regional server, the Group server calculates the differences between the data received from the Regional server and the data stored previously in its XOR_GlobalTable

- Updates XGBData_p_GB(t) (= XGBData_p_GB(t-1) $\oplus$ Packet_XRData_p_Rk(t) (for k $\in$ K(t)) and restore it in its XOR_GlobalTable

3 Conclusion and Future Research

In this paper, data synchronization operations for a hierarchical monitoring system to be applied when a very large number of PHD clients are used are proposed in order to distribute workload across the system. Detailed execution mechanisms for data synchronization operations (ADD, REPLACE) are described at every component in the system. After every component uses XOR operations to calculate the difference between what it had in the past and what it has now, it stores the difference in its own table and sends the difference to the component in the upper layer in order to reduce storage/communication costs.

Another reason to use XOR operations is that the operations can be used in developing fault recovery schemes for the system. Currently, a fault recovery scheme for the system using the XOR operations is being studied.

Acknowledgments. This work (Grants No. 0002985) was supported partly by Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2012. Also, this research was supported partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (No. 2012-013549).

References

1. IEEE Std 11073-10404 TM- 2008, Health Informatics-Personal Health Device Communication - Device Specialization-Pulse Oximeter (2008), `http://www.ieee.org`
2. IEEE Std 11073-10407 TM- 2008, Health Informatics-Personal Health Device Communication - Device Specialization-Blood Pressure Monitor (2008), `http://www.ieee.org`
3. IEEE Std 11073-10417 TM- 2008, Health Informatics-Personal Health Device Communication - Device Specialization-Glucose Meter (2008), `http://www.ieee.org`
4. IEEE Std 11073-10443 TM- 2008, Health Informatics-Personal Health Device Communication - Device Specialization-Physical Activity Monitor (2008), `http://www.ieee.org`
5. IEEE Std 11073-10472 TM- 2008, Health Informatics-Personal Health Device Communication - Device Specialization-Medication Monitor (2008), `http://www.ieee.org`
6. Park, K., Pak, J.: Efficient Transmission Method for Mobile Data Synchronization Based on Data Characteristics. In: Proceedings of the International Conference on IT Convergence and Security 2011, pp. 253–263 (2012)
7. Pak, J., Park, K.: UbiMMS: An ubiquitous medication monitoring system based on remote device management methods. The Journal of Healthcare Information Management 41, 26–30 (2012)
8. Pak, J., Park, K.: Design of an ISO/IEEE, 11073 gateway for u-healthcare services. In: The International Conference on Information Science and Technology, pp. 152–154 (2012)
9. Pak, J., Park, K.: Advanced pulse oximetry system for remote monitoring and management. BioMed Research International 2012 (2012)
10. Park, K., Pak, J.: An Integrated Gateway for Various PHDs in U-Healthcare Environments. BioMed Research International (2012)

Power Flow Analysis and Visualization in Smart Grid

Dan Wu[1], Rundong Wu[2], Zhijian Chen[1], Wenyan Xie[1], Xiang Huang[1], Bin Sheng[2,3,*], and Dawei Liu[4]

[1] Network & Information Branch, Guangdong Electric Power Design Institute, Guangzhou, China
[2] Dept. of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
shengbin@sjtu.edu.cn
[3] State Key Lab. of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China
[4] Dept. of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

Abstract. With the development of society, the requirement for security and economy in power grid is increasing. Hence, the concept of smart grid has been proposed. Smart grid is a new power network, which integrates the technology of advanced sensor measuring, network, communication, calculation and intelligent control. The application of visualization is a significant part in smart grid. On the other hand, the analysis of power flow in power grid is a traditional part in power system. This paper combines power flow analysis with visualization technology in one platform through C programming language. A detailed explanation is given on how to achieve power flow analysis through the classic Newton-Raphson algorithm and model the power grid with OpenGL programming method in computer. The platform can achieve the function of modeling a power system and dynamic effect in the line.

Keywords: Smart grid, power flow analysis, visualization, Newton-Raphson algorithm, OpenGL.

1 Introduction

Power flow Analysis plays a great role in the steady-state analysis and transient analysis. The main task of power flow is to calculate the voltage value, phase angle, power distribution and power loss of bus bars in power grid. Results of power flow can be applied into planning and design of power system. It can also provide basis and data for the selection of power grid's running mode and some protection for power grid. Visualization technology is widely used in many fields. It's used in modeling, simulation, monitoring, designing and control. The combination of visualization and power system promotes the security and economy of power grid greatly. [1]

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_60, © Springer-Verlag Berlin Heidelberg 2014

Plenty of research has been done in power flow analysis. However, the detailed programming method is deficient. There is a lack on how to compile the code of power flow and how to transform mathematical form into program form. The visualization technology has not been fully applied in power grid. At last, the visualization platform adopted nowadays is still abstract and not human-friendly. And the existing visualization platform is not intuitive, and not quick to access and obtain the useful information, especially when encountering with the fault. [2]

As a result, it's necessary to carry out some research on these aspects. This paper gives a detailed explanation on the achievement of power flow algorithm. We discuss how to program though Newton-Raphson algorithm and transform the Jacobian matrix from mathematical form to program form. Furthermore, the paper adopts OpenGL to model the power grid in computer and fulfill the dynamic effect of power in the line. Hence, it’s useful to program referring to power flow analysis quickly and observe the running condition of whole power grid directly and vividly.

2 Related Work

The initial power flow analysis is through manual calculation and AC & DC network calculator. Later, Gauss-Seidel method, Newton-Raphson method and Fast-Decoupled method is widely used in power flow analysis. [1] And Newton-Raphson method has the advantage of fast speed of convergence, yet influenced by the initials greatly. So the selection of initials is important for Newton-Raphson method and this method is often adopted in distribution system. [3] After that, Dommel and Tinney proposed Simplified Gradient method, which could meet the demand of optimization of power flow. Recently, more methods specialized at certain system is discovered and proposed, such as, Krylov subspace method for well-conditioned system and Broyden method for distribution system. [4, 5]

Visualization is a kind of technology which turns abstract objects or course to graphic or image expression. Visualization includes 4 branches: science computation visualization, data visualization, information visualization and knowledge visualization. [6] Furthermore, Power World Simulator is simulation software, developed by Power World Company, widely used all around the world. There are 3 main approaches to develop visualization product: VTK, Java 3D and OpenGL. The environment of OpenGL for 2D and 3D graphic program is the mostly applied computation graphic standard. [7]

3 Approach Overview

The main concept of Newton-Raphson algorithm is iteration. Assume the initial, set up and solve the Correction Equation. On the right of the equation, there is a Jacobian matrix. As there are four sorts of parameter in Jacobian matrix, four two-dimensional arrays are used to store each sort of parameter respectively. Calculate the parameter and store it into the arrays through loop structure. After that, the four arrays are stored into a one-dimensional array used for the solution of Correction Equation. Finally solve the Correction Equation and test the convergence. If it's not convergent, continue iterating. This is the brief overview of power flow analysis.

Next, we need to model the power grid and place the results of flow analysis in the power grid. Meanwhile, the power in the line can flow from one bus bar to another bus bar like water flowing in river. What's more, dispatchers can observe the figure in the line or equipment to identify the running condition of power system.

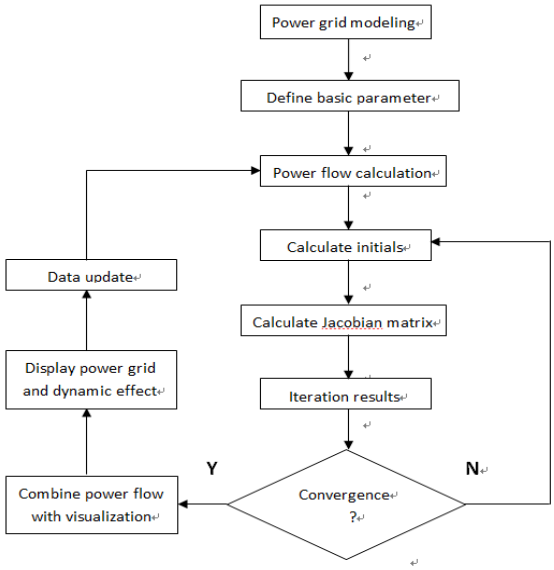


Fig. 1. Flow diagram for power system modeling

This is an overview of the whole power system modeling. The core part to achieve it in detail is discussed in the next two headings: power flow algorithm and visualization.

4 Power Flow Algorithm

In the analysis of power system, bus bars in power grid are always abstracted to power nodes, which indicate the voltage and phase angle in a certain place. Usually, the power nodes can be classified into three sorts. The first one is the PQ node, the second one is the PV node and the last one is the balancing node.

Correction Equation: The above power equations under the initial develop into a Taylor series. Omit the quadratic term and above of $\Delta\delta_i$ and $\Delta U_i/U_i$. The Correction Equation is as below,

$$\begin{pmatrix} \Delta P(1) \\ \Delta Q(1) \\ \cdots \\ \Delta P(m) \\ \Delta Q(m) \\ \Delta P(m+1) \\ \cdots \\ \Delta P(n-1) \end{pmatrix} = \begin{pmatrix} H(1,1) & N(1,1) & \cdots & H(1,m) & N(1,m) & H(1,m+1) & \cdots & H(1,n-1) \\ J(1,1) & L(1,1) & \cdots & J(1,m) & L(1,m) & J(1,m+1) & \cdots & J(1,n-1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ H(m,1) & N(m,1) & \cdots & H(m,m) & N(m,m) & H(m,m+1) & \cdots & H(m,n-1) \\ J(m,1) & L(m,1) & \cdots & J(m,m) & L(m,m) & J(m,m+1) & \cdots & J(m,n-1) \\ H(m+1,1) & N(m+1,1) & \cdots & H(m+1,m) & N(m+1,m) & H(m+1,m+1) & \cdots & H(m+1,n-1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ H(n-1,1) & N(n-1,1) & \cdots & H(n-1,m) & N(n-1,m) & H(n-1,m+1) & \cdots & H(n-1,n-1) \end{pmatrix} \begin{pmatrix} \Delta\delta(1) \\ \frac{\Delta U(1)}{U(1)} \\ \cdots \\ \Delta\delta(m) \\ \frac{\Delta U(m)}{U(m)} \\ \Delta\delta(m+1) \\ \cdots \\ \Delta\delta(n-1) \end{pmatrix} \quad (1)$$

The parameters in the Jacobian matrix are formulated as follows [8]

$$\mathrm{H}_{ij} = \frac{\partial \Delta \mathrm{P_i}}{\partial \delta_\mathrm{j}}, \qquad N_{ij} = \frac{\partial \Delta \mathrm{P_i}}{\partial \mathrm{U_j}} \mathrm{U_j} \quad (2)$$

$$\mathrm{J}_{ij} = \frac{\partial \Delta \mathrm{Q_i}}{\partial \delta_\mathrm{j}}, \qquad L_{ij} = \frac{\partial \Delta \mathrm{Q_i}}{\partial \mathrm{U_j}} \mathrm{U_j} \quad (3)$$

5 Visualization

Modeling of Power Grid: The power grid mainly consists of generator, bus bar, transformer, load and line. The paper discusses some way to model the equipment above. The circle is used to represent the generator, the rectangle stands for the transformer, the stripe represents the bus bar, the arrow expresses the load and the power

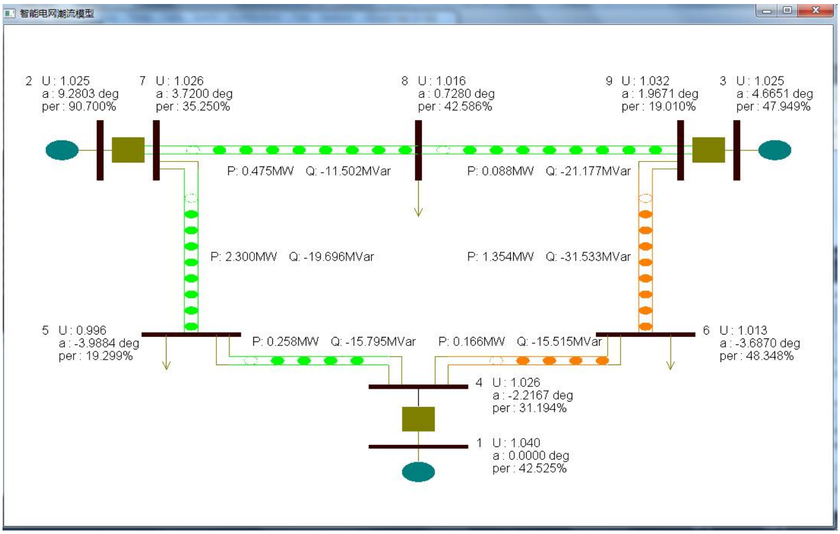


Fig. 2. The modeling power grid in example case

line is represented by a long straight line. For example, CreatLineHor(float pointX, float pointY, float length, float b, float p). CenterX and centerY(or pointX, pointY) represent the position of equipment in the computer screen. The b parameter shows the color of line, which indicates the running condition of power system. Color green means the power line is under light load. Color orange means it's overloaded. Color red means the power line is beyond the preset warning value, while the p parameter indicates the value of load percentage.

Achievement of Dynamic Effect: Dynamic effect is used to display the power flow in power system so that dispatchers can observe the direction and condition of power flow easily. The paper adopts the idle() function in OpenGL to achieve the dynamic effect. One is used for displaying the image. The other is used for drawing the image. After one frame is finished, the two buffering areas swap each other.

6 Experimental Results

Adopting the above ideas and methods, a simple platform of power grid is completed to verify the result of power flow algorithm and visualization. The power grid in the case is a system with 3 generators and 9 bus bars. The result data is as below,

Table 1. The result of example case

Bus bar	Voltage level/kV	Voltage value		Phase angle	
		Per unit	Known value/kV	Real value/rad	Angle value/deg
1	16.50	1.04	17.16	0.00	0.00
2	18.00	1.025	18.45	0.161972	9.2803
3	13.80	1.025	14.145	0.0814222	4.6651
4	230.00	1.02584	235.9432	-0.0386883	-2.2167
5	230.00	0.995669	229.00387	-0.0696113	-3.9884
6	230.00	1.01269	232.9187	-0.064351	-3.6870
7	230.00	1.02578	235.9294	0.064927	3.7200
8	230.00	1.01589	233.6547	0.0127053	0.7280

7 Conclusions

In this paper, we have proposed the method of modeling the power system. The platform can achieve the power flow analysis and display the power grid vividly. The Jacobian matrix in Correction Equation can be blocked by four small matrixes. Each small matrix can be solved independently, and compose the whole Jacobian matrix. The method of double buffering in OpenGL can be used to implement the dynamic effect. Modify the parameter, draft picture in one buffering, and display the picture in the other buffering.

Acknowledgements. This work was supported in part by the National Basic Research Project of China under Grant 2011CB302203, and in part by the National Natural Science Foundation of China under Grant 61202154 and Grant 61133009, in part by Shanghai Pujiang Program under Grant 13PJ1404500, in part by the Open Project Program of the National Laboratory of Pattern Recognition, and in part by the Open Project Program of the State Key Laboratory of CAD&CG, Zhejiang University, under Grant A1206.

References

1. Liu, G.: Module Development of Load Flow Calculation in Power Grid. Dissertation, Jilin University
2. Yan, Z., Sun, P., Liu, X., Zhou, W.: Design and Implementation of Visualization Platform in Power System. In: 2011 The International Conference Advanced Power System Automation and Protection, pp. 853–856. IEEE Press, New York (2011)
3. Wang, S., Liu, Y.: Review of Load Flow Calculation Methods in Power Systems. Shandong Electric Power 91, 8–12 (1996)
4. Saha, D., Singha, S.: A New Approach to Load Flow Analysis using Krylov Subspace Methods for well conditioned systems. In: 2011 Annual IEEE India Conference (INDICON), pp. 1–6 (2011)
5. Yang, H., Wen, F., Wang, L.: Newton-Raphson on Power Flow Algorithm and Broyden Method in the Distribution System. In: 2nd IEEE International Conference on Power and Energy (PECon 2008), pp. 1613–1618 (2008)
6. Yang, N., Yuan, S., et al.: Discussion on the Application of Visualization Technology in Modeling and Simulation. In: 2011 International Conference on Information Management, Innovation Management and Industrial Engineering, pp. 225–228 (2011)
7. Shen, G., She, D., Sun, P., et al.: Research and Application of Power System Visualization Technology. Power System Technology 33(17) (2009)
8. Ma, L., Zhang, B.: Analysis and Compare Newton-Raphson method to P-Q decoupled method on C++. The World of Power Supply 11, 32–35, 55 (2008)

Finite Difference Approach to Steady State Problems Arising from Mortgage and Option Pricing

Dejun Xie[1], Jin Zheng[2], Nan Zhang[3], Kun Chen[1], and Huaiqing Wang[1]

[1] Department of Financial Mathematics and Financial Engineering, South University of Science and Technology of China, China
{xie.dj,chen.k,wang.hq}@sustc.edu.cn

[2] Department of Mathematical Sciences, University of Liverpool, UK
Jin.Zheng@liverpool.ac.uk

[3] Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, China
nan.zhang@xjtlu.edu.cn

Abstract. Motivated by mortgage valuation, the paper proposes a finite difference approach to solve a class of free boundary problems which may be useful for option pricing in general. Given certain financially meaningful conditions, a mortgage borrower wishes to find the level of market interest rate at which it is optimal to make prepayment. The problem is an analog of finding the optimal level of stock price for early exercise in American put. Mathematically they both can be formulated as free boundary problems. In this paper an algorithm based on the finite difference scheme is designed to find the numerical solution to the steady state of such problems. The approach is calibrated with the perpetual American put option whose solution is explicitly known. The efficiency of the algorithm is tested by numerical simulations.

Keywords: Finite Difference, Numerical Methods, Mortgage Valuation, Option Pricing.

1 Formulation of the Problem

We consider a mortgage contract where the borrower pays a fixed rate of c $(year^{-1})$ to the lender. In reality this mortgage rate is implicitly represented by a continuous payment of m $(dollar/year$. At each time t when the contract is effect, the borrower has two choices: to continue the mortgage by paying mdt for the next dt period or to close the mortgage by paying off all the loan balance $M(t)$, where the loan balance $M(t)$ is determined by

$$\frac{dM(t)}{dt} = -m + cM(t). \tag{1}$$

When $M(T) = 0$ is specified for given $T > 0$, the above ODE has a unique solution

$$M(t) = \frac{m}{c}(1 - e^{(-c(T-t))}). \tag{2}$$

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_61, © Springer-Verlag Berlin Heidelberg 2014

Here we assume the borrower always has sufficient amount of capital. The borrower chooses not to pay $M(t)$ even though he is financially capable to do so if the expected future market return from an equal amount of investment is high enough. On the other hand, if the expected future market return from an equal amount of investment is lower enough, he should choose to settle $M(t)$. Assume that market interest follows the CIR model [2], from mathematical finance theory (see [5], [1] for instance), one can find the value of the contract $V(x,t)$ and the optimal level of market interest $x = h(t)$ at which the borrower should make prepayment of $M(t)$ by solving the following free boundary problem:

$$\begin{cases} \mathbf{L}(V) = m, & \text{for} \quad x > h(t), t > 0 \qquad (3) \\ V = \dfrac{m}{c}[1 - e^{-ct}], & \text{for} \quad x \le h(t), t > 0 \\ \dfrac{\partial V}{\partial x}(h(t), t) \equiv 0 & \qquad (4) \\ V(x, 0) = 0, & \text{for all} \quad x \ge 0 \\ h(0) = c \end{cases}$$

where t denotes the time to expiry of the contract, $h(t)$ is the unknown free boundary to be determined together with V, and the differential operator $\mathbf{L}$ is defined as

$$\mathbf{L}(V) = \frac{\partial V}{\partial t} - \frac{\sigma^2}{2} x \frac{\partial^2 V}{\partial x^2} - k(\theta - x)\frac{\partial V}{\partial x} + xV \tag{5}$$

Here we focus on the infinite horizon scenario where the duration of the contract is treated as infinite and the value of the contract does not vary in time. Thus we have the following equation that V, as a function in x, must satisfy.

$$xV'' + (\frac{2k\theta}{\sigma^2} - \frac{2k}{\sigma^2}x)V' - \frac{2}{\sigma^2}xV = -\frac{2c}{\sigma^2}, \qquad x \in (R^*,\ \infty) \tag{6}$$

where R^*, the lowest market rate at which the debtor chooses to settle all the loan balance, is to be determined together with V. To simplify the problem, we consider the discrete case of the infinite payment problem, where the debtor perpetually pays m dollar per year. The present value of the future payment is

$$PV = m[\frac{1}{1+R^*} + \frac{1}{(1+R^*)^2} + \frac{1}{(1+R^*)^3} + \cdots] = \frac{m}{R^*}$$

The problem is then described by the following system:

$$\begin{cases} xV'' + (\frac{2k\theta}{\sigma^2} - \frac{2k}{\sigma^2}x)V' - \frac{2}{\sigma^2}xV = -\frac{2c}{\sigma^2}, \qquad x \in (R^*,\ \infty) \\ V(R^*) = \frac{m}{R^*} \\ V_x(R^*) = -\frac{m}{(R^*)^2} \\ V(x = \infty) = 0 \end{cases} \tag{7}$$

A similar problem is formulated in [6], whereas the boundary conditions in [6] and the current study are different. Also, [6] does not concern numerical methods to the problem. The rest of the paper is organized as follows. In section two,we extend the above CIR model based differential equation to a generalized nonlinear second order non-homogeneous differential equation with similar boundary conditions at non-fixed boundary points. Finite difference method is applied to solve this system. Section three tests our generalized finite difference scheme with the perpetual American put option. Section four provides numerical results of the infinite mortgage problem using the tested algorithm. The paper is concluded with a wrap-up discussion.

2 Numerical Algorithm

Here we believe a bisection scheme is appropriate to find the numerical solution to the problem. As a semi-fixed boundary problem, the solution is given at the right end of the domain $[0, \infty)$. If we could possibly discretize the whole domain into a union of subintervals, then we may, start from the right end of the domain, apply a finite difference method to find the solutions at each time node in a backward direction. At each step, we may check the derivative of $\frac{\partial V}{\partial x}$ before moving onto next step of computing. The work loop stops when the derivative is zero (at given tolerance level). The major difficulty is that the original ODE is defined on an infinite domain $[0, \infty)$. Error will be introduced if we truncate the infinite domain into a finite one, say, $[0, \alpha)$, as a treatment frequently pursued in scientific computing. To address this infinite domain problem, we propose the change of variables

$$y = 1 - e^{-x},$$

which gives

$$\begin{cases} V''(1-y)^2 \ln(1-y) - (1-y)[(1+\dfrac{2k}{\sigma^2})\ln(1-y) + \dfrac{2k\theta}{\sigma^2}]V' - \dfrac{2}{\sigma^2}\ln(1-y)V \\ \qquad = \frac{2c}{\sigma^2}, \qquad y \in (y^*, 1) \\ V(y^*) = -\frac{m}{ln(1-y^*)} \\ V_y(y^*) = -\frac{1}{(1-y^*)(ln(1-y^*))^2} \\ V(1) = 0 \end{cases} \tag{8}$$

One can see that original infinite interval (R^*, ∞) is scaled into a finite interval $(y^*, 1)$. The benefit of such an easy change of variable not only comes from scaling the interval. More importantly, a uniformed mesh in the transformed system actually corresponds to an adaptive mesh in the original system. That is, as x approaches the boundary R^* from right side, the mesh size decreases exponentially. This surely increases the accuracy of the computation since the error of the finite difference scheme largely occurs at $x \to R^*$. To ease notations, we study the following general problem:

$$\begin{cases} pV'' + qV' + rV = g, \qquad y \in (y^*, 1) \\ V(y^*) = -\frac{m}{ln(1-y^*)} \\ V_y(y^*) = -\frac{1}{(1-y^*)(ln(1-y^*))^2} \\ V(y=1) = 0 \end{cases} \tag{9}$$

To solve this problem numerically, we appeal to finite difference method. Without loss of generality, let $\alpha = 1$. Partition the interval $[y^*, 1]$ into N equal length subintervals by points $y_1, y_2, ..., y_N$. Each subinterval has a length $dy = \frac{1-y^*}{N}$. Apply the central scheme of the difference equations to the first and second order derivatives, we have

$$V_{i+1}(\frac{p_i}{dy^2} + \frac{q_i}{2dy}) + V_i(\frac{-2p_i}{dy^2} + r_i) + V_{i-1}(\frac{p_i}{dy^2} - \frac{q_i}{2dy}) = g_i, i = 2, 3, ..., N$$

Since V_1 and V_{N+1} are known, the unknown variables are $V_2, V_3, ..., V_N, N-1$ unknown in total. Overall, we have such a linear algebraic system to solve

$$AV = B \tag{10}$$

where

$$V = \begin{pmatrix} V_2 \\ V_3 \\ \vdots \\ V_N \end{pmatrix} \tag{11}$$

$$A = \begin{pmatrix} \frac{-2p_2}{dy^2} + r_2 & \frac{p_2}{dy^2} + \frac{q_2}{2dy} & & \\ \frac{p_3}{dy^2} - \frac{q_3}{2dy} & \frac{-2p_3}{dy^2} + r_3 & \frac{p_3}{dy^2} + \frac{q_3}{2dy} & \\ & & \ddots & \\ & & & \frac{p_N}{dy^2} - \frac{q_N}{2dy} & \frac{-2p_N}{dy^2} + r_N \end{pmatrix} \tag{12}$$

$$B = \begin{pmatrix} g_2 - V_1(\frac{p_2}{dy^2} - \frac{q_2}{2dy}) \\ g_2 \\ \vdots \\ g_{N-1} \\ g_N - V_{N+1}(\frac{p_N}{dy^2} + \frac{q_N}{2dy}) \end{pmatrix} = \begin{pmatrix} g_2 - (\frac{p_2}{dy^2} - \frac{q_2}{2dy}) \\ g_2 \\ \vdots \\ g_{N-1} \\ g_N \end{pmatrix}$$

This linear system is solvable when y^* is prescribed. We remark that the stability condition is automatically satisfied for $y^* > 0$. In our problem, p, q, r are determined by the following

$$p_i = (1-y_i)^2 \log(1-y_i) \tag{13}$$

$$q_i = -(1-y_i)[(1+\frac{2k}{\sigma^2})\log(1-y_i) + \frac{2k\theta}{\sigma^2})] \tag{14}$$

$$r_i = -\frac{2}{\sigma^2}\log(1-y_i) \tag{15}$$

for $i = 2, 3, ..., N$. Now the problem becomes to find a suitable y^* to meet the first derivative condition at optimal exercise boundary. Thus we propose the following bisection scheme. For given parameter values, we start with a suitable initial guess of an interval covering y^*. Practically, $[0, c]$ is large enough. After each iteration of bisection, $|M_1 - M_2|/M_1$ is recorded. In our experiment, we compare the first differentiation of V_y at y^*, with the result obtained by finite difference method. Hence, M_1 is the theoretical $V_y(y^*)$, where M_2 is the real slope of the function V at the point y^*. The loop stops when a prescribed lowest error tolerance is reached. A suedo-Matlab program is as follows.

1. Define k, θ, σ, c, where k, θ, σ are parameters for the CIR model, c is the contracted mortgage rate, and N is the number of subintervals in x direction.
2. Define vectors p, q, r, each has N entries. And define the matrix A whose entries are given by equation (12).
3. For an initial guess of an interval covering y^*, say $[y^{*1}, y^{*2}]$, find the solution vectors V by solving the linear system. Record relative error= $|M_1 - M_2|/M_1$.
4. Let $y^{*3} = \frac{1}{2}(y^{*1} + y^{*2})$. Find the corresponding solution vector and record relative error= $|M_1 - M_2|/M_1$. Update y^{*2} with either y^{*1} or y^{*2}. Repeat the above procedures until the minimum relative error is achieved.

One may also consider the power series method to solve the linear ODE with variable coefficients. However, note that the coefficient of V'' has a singular point zero at $x = 0$, and it will affect the existence of the power series. Hence, a numerical method is more realistic for solving the problem.

3 Implementation

The robustness of our method can be examined by special cases of (9). In particular, we test it with the perpetual American put, the solution of which is explicitly known. Consider the ODE and the boundaries of perpetual American put:

$$\begin{cases} \frac{1}{2}\sigma^2 y^2 \frac{\partial^2 V}{\partial y^2} + Ry\frac{\partial V}{\partial y} - RV = 0 \\ V(y^*) = k - y^* \\ V'(y^*) = -1 \\ V(\infty) = 0 \end{cases} \tag{16}$$

The corresponding finite difference scheme for the problem is

$$V_{i+1}(\frac{p_i}{dy^2} + \frac{q_i}{2dy}) + V_i(\frac{-2p_i}{dy^2} + r_i) + V_{i-1}(\frac{p_i}{dy^2} - \frac{q_i}{2dy}) = g_i, i = 2, 3, ..., N$$

where $p_i = \frac{1}{2}\sigma^2 y_i^2, q_i = Ry_i, r_i = -R, q_i = 0$. The solution to the free boundary y^* is determined by the condition

$$\frac{\partial S}{\partial y}|_{y=y^*} = -1$$

The accuracy of the algorithm can be tested against the explicit solution known as

$$y^* = \frac{2kR}{2R + \sigma^2}$$

The value of the contract can be then computed as

$$V_{y^*} = (k - y^*)(\frac{y}{y^*})^{-\frac{2R}{\sigma^2}}$$

The following Figure 2 compares the numerical solution to the perpetual American put option using our algorithm and the corresponding true solution, where typical option parameters are tested with $y_{max} = 2000$. For fixed option parameters, the error between the two methods diminishes to zero as the mesh size is sufficiently refined. To numerically test our algorithm for solving the infinite mortgage payment problem, we set the contractual interest rate $r_0 = 0.056$. and the yearly payment $m = 1$. The following Figure 1 provides the numerical results for the case where k=0.1 and θ=0.05. Outputs for different values of σ are compared.

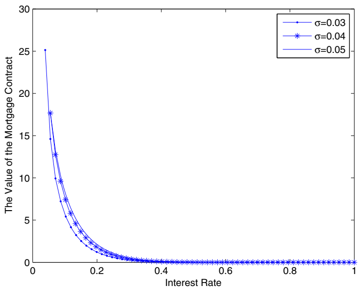


Fig. 1. Value of mortgage for different σ

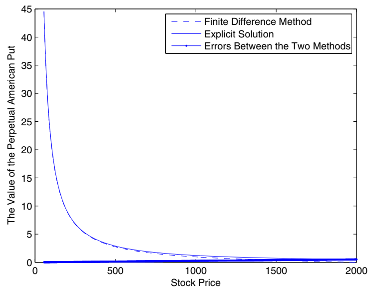


Fig. 2. Value of perpetual American put when $y_{max} = 2000$

4 Concluding Remark

The current work focuses on finding the best strategy for valuing the mortgage contract in the infinite horizon environment. The optimal rate for prepayment is obtained by iterative finite difference scheme. The finite difference method is generalized to describe the governing ODEs for a broad class of option pricing problems. The effectiveness of this method has been verified by the perpetual American put. As a future direction, further theoretical evaluation of the method including convergence analysis can be conducted.

References

1. Baxter, M., Rennie, A.: Financial Calculus, An introductin to derivative pricing. Cambridge University Press (2005)
2. Cox, J., Ingersoll, J., Ross, S.: A theory of the term structure of interest rates. Econometrica 53, 385–407 (1985)
3. Vasicek, O.A.: An equilibrium characterization of the term structure. J. Fin. Econ. 5, 177–188 (1977)

4. Friedman, A.: Variational Principles and Free Boundary Problems. John Wiley & Sons, Inc., New York (1982)
5. Wilmott, P., Howison, S., Dewynne, J.: The Mathemaitcs of Financial Derivatives. Cambridge University Press (2002)
6. Xie, D.: An Steady State Problem Arising from Mortgage Valuations, arxiv.org/abs/0909.5389 (2009)
7. Ascher, U.M., Petzold, L.R.: Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. Society for Industrial and Applied Mathematics (1998)

A Design for Commonsense Knowledge Enhanced High-Frequency Market Analysis System

Kun Chen[1], Huaiqing Wang[1], Jian Yin[2], and Dejun Xie[1]

[1] Department of Financial Mathematics & Engineering, South University of Science and Technology, 1088 Xueyuan Blvd., Shenzhen, Guandong, China, 518055
{chen.k,wang.hq,xie.dj}@sustc.edu.cn
[2] Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong
jyin@eee.hku.hk

Abstract. To identify the impacts of public news on security market, we propose a common-sense knowledge supported news analysis method, and design a system architecture for the news incorporated market analysis system. The graph model of common-sense knowledge is used to extend the news feature set by a random walk method. News indicators including news sentiment and news relevance are measured by common-sense knowledge supported text mining techniques. Based on these ideas, we develop a prototype system and examine the intra-day market reactions to public news on Hong Kong stock market. Our finds have shown the effectiveness of using common-sense knowledge and news in the market analysis domain. It is our belief that the common-sense knowledge incorporated market analysis system would be great helpful to market surveillance bureaus, traders and investors on the security market.

Keywords: commonsense knowledge, market analysis, system design.

1 Introduction

In modern financial markets, information is disseminated across the globe almost instantaneously, a situation which requires competitive institutions to analyze and react to market events at high speed. Trading on financial market is ultimately driven by news. However, news is difficult to identify and to analyze in a real-time way. In particular, it is problematic to identify the relevant news and to interpret them accordingly, considering the huge amount of information available today. Because of these difficulties, nearly all empirical studies examine the impact of news by solely focusing on specific news events, such as scheduled macroeconomic announcements, political interventions, or certain firm-specific news such as earning announcements which are in most cases easily identifiable. Given the role that common-sense knowledge plays in the human's cognitive process, i.e., connecting different concepts, arguments, facts, and events to create an inference model and employing semantic relations between concepts to decode information hidden in texts, we design an innovate model to measure market news sentiment and relevance. A common-sense

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_62, © Springer-Verlag Berlin Heidelberg 2014

knowledge supported market analysis system is designed to incorporate common-sense knowledge and public news in the market analysis process.

The remainder of the paper is organized as follows. In the next section, we describe the background research on news analysis and common-sense knowledge. Section 3 presents the common-sense knowledge based news analysis method. A prototype system is developed and evaluated in section 4 and section 5. Section 6 concludes.

2 Background Research

The question of how news is incorporated into asset prices is analyzed by a wide range of studies. However, most of these literatures focus on macroeconomic news and company-specific earnings announcements, and only very few studies try to link trading activities with an intraday flow of information. The predominant part of previous research investigates the change of returns and volatility, responding to the macroeconomic news by the event study method. The measured financial products include index futures, , interest futures , Forex futures. Considering the continuous intraday news flow, Grob-Klubmann and Hautsch [1] is the only study to combine the news sentiment and relevance as indicators to measure the 20 seconds high-frequency reactions of 40 stocks from the FTSE 100 Index to real-time news announcements.

Commonsense knowledge refers to the general world knowledge that an ordinary person should know. In general, commonsense knowledge can be represented as structured, semi-structured, or non-structured forms [2, 3]. The semi-structured representation, the most commonly used form in computer systems, describes commonsense concepts, facts, and their relations in a graph structure . In such ontology, each node represents a concept or an entity (i.e., an instance of the concept) and each link represents the commonsense relations between concepts/entities. Aligned with these representations, significant effort has focused on organizing human commonsense knowledge into knowledge bases [4, 5]. One type of knowledge base focuses on general concepts and rules of the world in a strict formalization. The Cyc knowledge base [5] started in 1984 by Cycorp is a famous example, which now contains hundreds of thousands of facts and millions of relations represented by a specific formal language CycL. A second type of commonsense knowledge base focuses on documenting concrete entities and instantiations of conceptual thing, such as DBpedia [6] and YAGO [7].

3 Common-Sense Knowledge Incorporated News Analysis Model

The common-sense knowledge is organized in a graph structure. The issuer companies of stocks are linked with related news in the framework based on common-sense knowledge. Specifically, each security is mapped with an issuer company, and each news article is represented by a group of features. Both the issuer company and news features are mapped with the concepts/entities in the common-sense model. By the links/properties defined in the conceptual model, the issuer company could be linked with news features. In this way, the market activities about the particular security are related with the news that contains a feature linked with its issuer company. We identify the news relevance relations by this mapping mechanism.

On the other hand, the news features are inter-linked with each other according to the relations defined in the conceptual model. Such relations can be used for semantic smoothing and feature generation within news articles. For example, the system identifies that the words of "good" and "well" have the same meaning in common-sense knowledge base, it may replace them with each other in the feature set. This kind of concept-words relations can also be identified by walking several steps in the common-sense conceptual model.

The graphical conceptual model provides the news relevance and news semantic relations in this way to ease human reasoning. It can address the meta-requirement of our design product on supporting deep understanding of textual information and providing effective presentation of the identified cues for decision making on the financial market.

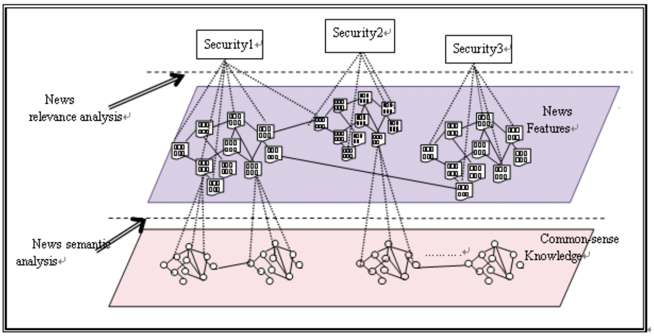


Fig. 1. A framework for common-sense knowledge incorporated market analysis system

4 System Implementation

Following our proposed method, we built a prototype system, CS-MAS, targeted at analyzing the high-frequency market reactions in Hong Kong stock market. In this research, we focus on 42 Hang Seng Index (HSI) component stocks in the sample period from Jan 2, 2009 to Dec 31, 2009.

We gather the public real-time news from 3 online news agencies. The first one is "Bloomberg.com", which posts the world wide breaking news at real time. The second one is "The Standard", which is a Hong Kong local news agency that provides most local news. The last one is the "HKEx news" website, which publishes the announcements and notices disclosed by the listed companies. If a piece of news appears or is updated several times, we just keep the earliest one. We also delete the over-night news, say, from the market close time of the current trading day (16:00) to the market open time (10:00) of the next trading day. To facilitate the analyzing of short term market reactions to the public released breaking news, we use 5 minutes as the time interval followed by the previous research [8,9,10].

In this system, we incorporated three common-sense knowledge bases. The first one is OpenCyc (v2.0), which is the open source version of the Cyc project. It contains about 150K terms/concepts and about 1,430K relations. The second one is DBpedia (v3.5), which contains structured information extracted from Wikipedia, including about 3.4 million entities (312K persons, 413K places, 140K organizations, etc.) and about 1 billion relations. The third knowledge base is the Senti-word Net (v1.0) (SentiWordNet 2009), which is to provide the sentiment score of common used words/phrases.

5 System Operation

Most daily news is announced during the market off period at noon in Hong Kong, and followed by the opening time of market. For the company news, we identify 192 time intervals and accumulate 279 news articles yearly. For the common-sense extended first level company related news, there are 1893 time intervals and collect 3514 news. For the common-sense extended second level company related news, 1904 time intervals are selected with 3486 pieces of news.

5.1 The Market Description

Figure 2 shows the intra-day trends in returns, return volatility, price volatility, trade volume, trade count, and traded money value for the 42 HSI stocks. The returns don't show strong daily patterns as measured by high-frequency data. Return volatility and price volatility show almost the same shape in the graph. They are rather flat within a day's continuous trading time and do not show an obvious U-shaped pattern, which is found in most previous work. However, the volatility increases sharply at the beginning

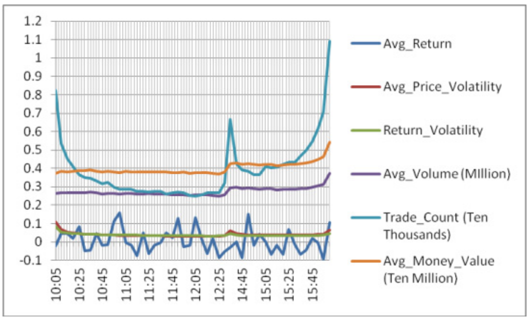


Fig. 2. Intra-day trends in returns, return volatility, price volatility, trade volume, trade count, and traded money value for the 42 HSI stocks

and the end of the day, as well as at the beginning of the afternoon trading section. This result may be explained by that informed trading increases in the three corresponding time periods as more information is released. The trade volume and traded money value show almost the same shape. No U-shaped pattern is found, but the lowest point is found in the end of morning trading, and the highest point is reached in the end of the day. The trade count shows a double U-shaped pattern in the morning and afternoon trading periods. This may be caused by the noon market off mechanism implemented on HKEx market.

We use Cyc and DBpedia as the common-sense knowledge in experiments. The numbers of news by different measure is shown in Figure 3. Most daily news is announced during the market off period at noon in Hong Kong, and followed by the opening time of market. Especially, for the announcements published on HKEx website, companies are more likely to avoid releasing information in continuous trading period.

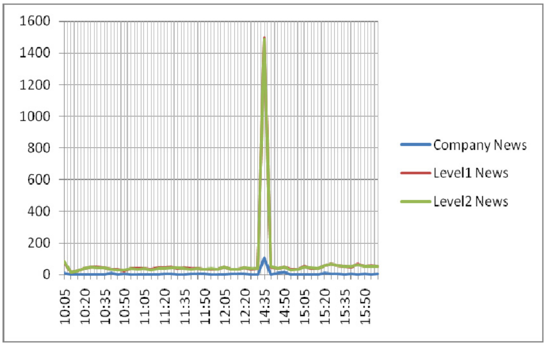


Fig. 3. The intra-day news distribution about the 42 HSI stocks collected by company name and common-sense knowledge

5.2 The News Impact on Market Performance

We firstly measure the differences of market performances before and after each news period. Table 1 shows the p-values of pared-t tests for 13 variables on the moving averages of 3-intervals lead and lag the identified news periods. If just using the company name as keywords to search and gather news, we notice that only the money value and spread are significantly different before and after the news periods. When extending the news searching range and combine the direct common-sense properties to gather news, we find the volume, trade count, money value, spread and depth are significantly different before and after the corresponding periods. If we further extend the news range, and use the indirect common-sense properties to collect news, all variables except the mid-quote price are significantly different between lead and lag periods.

Table 1. Comparison of trading activities and market liquidities before and after the identified news period

	Company	Level1	Level2
Avg_Return	0.7342	0.3983	0.0018
Return_Volatility	0.5356	0.0152	<.0001
Price_Volatility	0.2455	0.905	0.0002
Avg_Volume	0.2478	<.0001	<.0001
Volume	0.9125	<.0001	<.0001
Trade_Count	0.1101	<.0001	<.0001
Money_Value	0.0024	<.0001	<.0001
Avg_Money_Value	0.0004	<.0001	<.0001
Abs_Spread	0.0772	<.0001	<.0001
Mid_Price	0.4431	0.7457	0.4314
Relative_Spread	0.0214	<.0001	<.0001
Total_Ddepth	0.5258	<.0001	<.0001
Max_Ddepth	0.3703	0.0002	<.0001
Total_Vdepth	0.3194	0.0298	<.0001
Max_Vdepth	0.8464	0.0081	<.0001

6 Conclusions

Motivated by the ongoing surge in the amount of public news available on the stock market, this study focuses on a design of common-sense knowledge supported market analysis system. Our basic idea is to use common-sense knowledge to identify relations between news and stocks, as well as the relations between words and concepts. The system operation shows that, firstly, the common-sense knowledge and public news play an important role on security market, and should be considered when simulating the market; Secondly, our design can incorporate public news and other market measurements in quantitative market analysis models by using common-sense knowledge and text mining techniques; Thirdly, the design would be useful in many fields on security market, such as trading, market surveillance, and risk management and etc..

References

1. Grob-Klubmann, A., Hautsch, N.: Quantifying high-frenquency market reactions to real-time news sentiment announcements. SFB 649 Discussion Paper 2009-063 (2009)
2. Chen, K., Wang, H.Q., Lai, H.: A General Knowledge Mediation Infrastructure for Multi-agent Systems. Expert Systems with Applications 38(2011), 495–503 (2011)
3. Chen, H.C., Lynch, K.J.: Automatic Construction of Networks of Concepts Characterizing Document Databases. IEEE Transactions on Systems Man and Cybernetics 22(5), 885–902 (1992)

4. Storey, V.C., Sugumaran, V., et al.: A Semi-automatic Approach to Extracting Common Sense Knowledge from Knowledge Sources. Natural Language Processing and Information Systems, Proceedings 3513, 322–332 (2005)
5. Lenat, D.B., Guha, R.V.: Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project. Addison-Wesley Longman Publishing Co., Inc., Boston (1989)
6. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: Aberer, K., et al. (eds.) ASWC/ISWC 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007)
7. Suchanek, F.M., Gjergji, K., Gerhard, W.: Yago: A Large Ontology from Wikipedia and WordNet. Journal of Web Semantics 6(3), 203–217 (2008)
8. De Jong, F.A.M.W.M.D.: Intraday lead-lag relationships between the futures, options and stock market. European Finance Review 1, 337–359 (1998)
9. Stephan, J.A.A.R.E.W.: Intraday price change and trading volume relationships in the stock and stock options markets. Journal of Finance 45(1), 191–220 (1990)
10. Andersen, T.G., Boilers'let, T.: Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 42(2-3), 115–158 (1997)

Security Analysis of a Delegation-Based Authentication Protocol for Wireless Roaming Service*

Mijin Kim[1], Namje Park[2], and Dongho Won[1,**]

[1] College of Information and Communication Engineering
Sungkyunkwan University, Korea
{mjkim,dhwon}@security.re.kr
[2] Department of Computer Education Teachers College
Jeju National University, Korea
namjepark@jejunu.ac.kr

Abstract. Portable devices are widely used in our daily life. A delegation-based authentication is used for providing security and privacy for portable communication systems. In the current work, we are concerned with the security for a delegation-based authentication protocol. Firstly, we disclose vulnerability of Tsai et al.'s delegation-based authentication protocol for portable communication systems, then present a feasible solution that eliminates the potential threat to ensure the authenticity, integrity, and confidentiality of involved communications.

Keywords: Portable communication system, proxy signature, wireless roaming service, mobile authentication, known key security.

1 Introduction

Secure communication systems among network enabled devices are of significant concern in mobile access. This is often achieved by having the parties run an authentication protocol for generating a mutual and secret session key. The wireless roaming service allows a Visitor Location Register (VLR) to authenticate a visiting Mobile Station (MS), with the help of its Home Location Register (HLR). Mobile authentication among three parties may include sophisticated attacks, and is easily susceptible to masquerading [1, 2]. A secure communication system should provide secrecy, authenticity, integrity, and nonrepudiation features. If the system performance is not a main concern, RSA would be appropriate; otherwise, one needs to consider different alternatives. In 2005, Lee-Yeh introduced the concept of delegation to achieve security services with reduced cost in the wireless communication environment [3]. Delegation based protocol (We term this DBA protocol in short) is inspired by the proxy

* This research was supported by the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2013-H0301-13-3007) supervised by the NIPA(National IT Industry Promotion Agency).

** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_63, © Springer-Verlag Berlin Heidelberg 2014

signature, which is the delegation of the power to sign messages [4, 5]. The proxy signature is an authorized signature technique. The assistant is authorized to sign the document when the manager is away, but the staff can still use the manager's public key to verify the document. The manager cannot deny the signature if a dispute arises. Public key based system can benefit by the nonrepudiation feature of the public key cryptosystem. Each MS gets a different pair key (σ, K) from HLR in registration phase. The key implies the authorization from HLR. This authorization makes VLR transfer his trust in HLR to the requested legal pseudonym MS. HLR has the ability to identify the misused MS [3]. The delegation model does not increase computational loads for mobile stations, it provides good communication efficiency, data security, and user privacy. The DBA protocol to provide solutions to the problem of portable communication system. The significant advantage of use of trust delegation on mobile authentication is that a scheme can exploit the public-key based strong security properties while achieving efficiency in communication and computation through the use of a single symmetric key. For example, a MS in such a scheme does not require to have its own private key, hence there is no incurred security complication and overhead on public-key certificate of MS distribution which is particularly costly in a mobile communication environment. In the scheme, via delegation, a MS shares a symmetric key with its HLR. A MS in the scheme first signs a message in a similar fashion as that for delegation and sends it to a VLR so that a VLR can verify the validity of the delegation based on a public certificate published by HLR for this MS. Hence the VLR is able to authenticate a MS. After the verification, the VLR forwards the service request to the HLR. HLR can then forward the communication key to VLR after the shared secret is verified and VLR is authenticated. In 2012, Tsai et al. proposed a delegation-based authentication protocol for Wireless Roaming Service [6]. The protocol does not require proxy key pair synchronization between the HLR database and MS SIM card. Moreover, the computational cost of the Tsai et al.'s protocol is lower than existing protocols. In the Tsai et al.'s protocol, the session key SK is derived by three random numbers N_1, n_2, n_3 which are generated by MS, VLR, and HLR, respectively. In addition, MS has the ability to verify whether MS has correctly logged-in the current VLR by checking ID_V. What we do in this work is to report security vulnerability of Tsai et al.'s DBA protocol and show how to eliminate them.

2 Review of Tsai et al.'s Protocol

This section reviews the DBA protocol proposed by Tsai et al. [6]. The protocol consists of three phases: setup, online authentication, and *i*-th offline authentication. The following notations and acronyms are used through this paper.

- HLR: Home Location Register
- VLR: Visitor Location Register
- MS: Mobile Station

- $(x, v), (x_v, y_v)$: two private/public key pairs of HLR
- *p, q*: the prime numbers satisfying $q \mid (p-1)$
- *g*: a generator in
- ID_V, ID_H : the identities of VLR and HLR
- (σ, K) : a proxy key pair authorized by HLR
- $[M_K]$: a symmetric encryption for message M with the key K
- *G*: a cyclic additive group
- *h*(): an one-way hash function such that $h(): Z_p \to Z_p$
- *H*(): an one-way hash function such that $H(): \mathrm{G} \to Z_p$
- *P*: the generator of the cyclic additive group
- *n*: the total amount of time for offline authentication supported by the protocol
- n_1, n_2, n_3, k, t : random numbers in Z_p
- K_{HV} : the secret key between HLR and VLR
- *SK*: the session key between HLR and MS
- C_i : the session key between VLR and MS

2.1 Setup Phase Consisting of MS, VLR, and HLR

The setup phase includes the establishing of private and public keys, and proxy key pair for the DBA protocol.

Step 1. HLR chooses two private keys x and x_v, and then computes their corresponding public keys $v = xP$ and $y_v = x_v P$, respectively. Then, HLR shares $K_{HV}, x_v,$ and v with VLR. HLR also computes the proxy key pair $K = kP$ and $\sigma = x + kh(K) (\bmod q)$ for each MS.

Step 2. Each MS's generated proxy key pair is stored in HLR's database, and each MS's proxy key pair and public key are stored in each corresponding MSs SIM card.

2.2 Online Authentication Phase for MS, VLR, and HLR

For each online authentication session, MS chooses an n_1, generates a hash chain $h(n_1), h^2(n_1), \ldots, h^{(n+1)}(n_1)$, and stores them in its database.

Step 1. MS sends a login request to VLR.
Step 2. VLR selects a random number $n_2 \in Z_p,$ and sends (n_2, ID_v) to MS.

Step 3. MS retrieves $N_1 = h^{(n+1)}(n_1)$ from its SIM card, and computes $r_1 = tP, r_2 = h(ty_v) \oplus (K, N_1)$, and $s = \sigma \times h(N_1 \parallel n_2 \parallel ID_v \parallel r_1 \parallel r_2 \parallel ID_H) + t (\mathrm{mod}\ q)$. Then, MS sends $(r_1, r_2, s, ID_H, ID_V)$ to VLR.

Step 4. VLR uses x_v to retrieve K and N_1 by computing $r_2 \oplus H(x_v r_1)$. Then, VLR computes sP and $\alpha = h(N_1 \parallel n_2 \parallel ID_v \parallel r_1 \parallel r_2 \parallel ID_H) \times (v + h(K)K) + r_1,$ then verifies that sP is equal to α. If the verification fails, then VLR denies the login request. Otherwise, VLR computes $CT_1 = [N_1 \| n_2 \| K]_{K_{HV}}$ by using K_{HV} as the encryption key, and sends (CT_1, ID_H, ID_V) to HLR.

Step 5. Upon receiving HLR obtains $N_1 \| n_2 \| K$ by decrypting CT_1 with the secret key K_{HV}. Next, HLR finds its corresponding σ from its database according to the decrypted K. HLR then computes the session key $SK = h(N_1 \| n_2 \| n_3 \| \sigma)$, $CT_2 = [N_1 \| n_3 \| ID_v]_\sigma, CT_3 = [CT_2 \| n_2 \| N_1 \| SK]_{K_{HV}}$. Finally, HLR sends (CT_3, ID_H, ID_v) to VLR.

Step 6. After having received (CT_3, ID_H, ID_V), VLR obtains $CT_2 \| n_2 \| N_1 \| SK$ by decrypting CT_3 with the secret key K_{HV} and verifies whether n_2 and N_1 exist in the decrypted string $CT_2 \| n_2 \| N_1 \| SK$. If the verification holds, VLR sends (CT_2, ID_V) to MS.

Step 7. MS obtains $N_1 \| n_3 \| ID_V$ by decrypting CT_2 with the key σ in the MSs SIM card and checks whether N_1 and ID_V exist in the decrypted string $N_1 \| n_3 \| ID_V$. If the condition holds, MS computes the session key $SK = h(N_1 \| n_2 \| n_3 \| \sigma)$.

2.3 *i*-th Offline Authentication Phase for MS and the Current VLR

The offline authentication processes have been performed until i is equal to n, and then the online authentication process should be started again, if MS demands another authentication request. MS retrieves $h^{(n-i+1)}(n_1)$ from its database, and encrypts $h^{(n-i+1)}(n_1)$ with C_i, then sends $[h^{(n-i+1)}(n_1)]_{C_i}$ to VLR. Upon receiving $[h^{(n-i+1)}(n_1)]_{C_i}$, VLR decrypts the message and computes $h(h^{(n-i+1)}(n_1))$. Next, VLR verifies whether the computed value $h(h^{(n-i+1)}(n_1))$ is the same as the stored value $h^{(n-i+2)}(n_1)$ in its database. If the condition holds, VLR replaces $h^{(n-i+2)}(n_1)$ with $h^{(n-i+1)}(n_1)$, and computes the session key $C_{i+1} = h(h^{(n-i+1)}(n_1), C_i)$ and increases $i=i+1$.

3 Vulnerability of Tsai et al.'s Protocol

In this section, we reveal an violation of known key security on the DBA protocol described in Section 2 which cannot be detected by MS and HLR.

Violation of Known Key Security. In the *i*-th offline authentication phase, if the session key C_i is revealed to an adversary A, the session key compromise allows an active adversary to compromise keys of other sessions. The attack proceeds as follows:

We assume that the session key C_i is revealed to an adversary *A*. Then, *A* is able to decrypt $[h^{(n-i+1)}(n_1)]_{C_i}$ and obtain $h^{(n-i+1)}(n_1)$. Then, A is able to compute the next session key $C_{i+1} = h(h^{(n-i+1)}(n_1), C_i)$ by using C_i and $h^{(n-i+1)}(n_1)$.

This demonstrates that the DBA protocol proposed by Tsai et al. fails to provide known key security which is a fundamental requirement for secure communication.

4 Improved Protocol

4.1 Description of the Improved Protocol

The security vulnerability of the DBA protocol for wireless roaming service described in Section 3 is attributed to the following flaw in its design:

In the *i*-th offline authentication phase, VLR computes the *i*-th session key $C_{i+1} = h(h^{(n-i+1)}(n_1),\ C_i)$. Having identified the source of the problem, it is apparent how to repair the DBA protocol. We recommend the following changes to the DBA protocol.

MS | VLR

$\boxed{i\text{-}th \text{ offline authentication phase}}$

Retrieve $h^{n-i+1}(n_1)$ from its DB

$$\xrightarrow{[h^{n-i+1}(n_1), [h^{n-i-1}(n_1)]_{y_v}]_{C_i}}$$

Decrypt $[h^{n-i+1}(n_1), [h^{n-i-1}(n_1)]_{y_v}]_{C_i}$
Obtain $h^{n-i+1}(n_1)$
Decrypt $[h^{n-i-1}(n_1)]_{y_v}$
Verify $h^{n-i+1}(n_1) = h(h(h^{n-i-1}(n_1)))$
Replace $h^{n-i-1}(n_1)$ with $h^{n-i+1}(n_1)$
Compute $C_{i+1} = h(h^{n-i-1}(n_1), C_i)$
Increase $i = i + 1$

Fig. 1. The *i*-th authentication phase of the improved protocol

In the i-th offline authentication phase, MS sends $\beta = [h^{(n-i+1)}(n_1), [h^{(n-i-1)}(n_1)]_{y_v}]_{C_i}$ instead of $[h^{(n-i+1)}(n_1)]_{C_i}$ to VLR, then VLR decrypts β and obtains $h^{(n-i+1)}(n_1)$ and $[h^{(n-i-1)}(n_1)]_{y_v}$. Next, VLR decrypts $[h^{(n-i-1)}(n_1)]_{y_v}$ with private key x_v, then verifies $h^{(n-i+1)}(n_1) = h(h(h^{(n-i-1)}(n_1)))$. If the condition holds, VLR replaces $h^{(n-i-1)}(n_1)$ with $h^{(n-i+1)}(n_1)$ and computes the session key $C_{i+1} = h(h^{(n-i-1)}(n_1), C_i)$.

4.2 Security and Performance Discussion of the Improved Protocol

On the security point of view, these modifications described in Section 4.1 effectively prevent the vulnerability described in Section 3. The attack against known key secrecy can no longer be applied because, A can compute neither the token $h^{(n-i-1)}(n_1)$ nor $h^{(n-i)}(n_1)$ from $h^{(n-i+1)}(n_1)$ due to the hash function's one-way characteristics. Thus, an adversary can no longer generate C_{i+1} even with $h^{(n-i+1)}(n_1)$ at hand. Therefore, the adversary to attain the authentication secrets cannot succeed in impersonating at the next login. In the improved protocol, symmetric encryption $[h^{(n-i-1)}(n_1)]_{y_v}$ and decryption are additionally included for calculating the next session key between VLR and MS. However, the cost of symmetric encryption/decryption is negligible, the performance of the improved protocol is almost identical with that of Tsai et al.'s protocol.

5 Conclusion

This paper provides an overview of the vulnerability introduced by an attack, as well as the countermeasures to mitigate the threat. Our analysis demonstrates that Tsai et al.'s protocol does not provide known key security, which is a fundamental requirement for secure communication. Future work is under taken to present a formal proof.

References

1. Rhee, K., Won, D., Jang, S., Chae, S., Park, S.: Threat modeling of a mobile device management system for secure smart work. Electronic Commerce Research 13(3), 243–256 (2013)
2. Jeon, W., Kim, J., Nam, J., Lee, Y., Won, D.: An Enhanced Secure Authentication Scheme with Anonymity for Wireless Environments. IEICE Transactions on Communications E95-B(7), 2505–2508 (2012)
3. Lee, W., Yeh, C.: A New Delegation-Based Authentication Protocol for Use in Portable Communication Systems. IEEE Transactions on Wireless Communications 4(1), 57–64 (2005)
4. Lee, W.B., Chang, C.Y.: Efficient proxy-protected proxy signature scheme based on discrete logarithm. In: Proc. 10th Conference of Information Security, pp. 4–7 (2000)
5. Mambo, M., Usuda, K., Okamoto, E.: Delegation of the power to sign messages. IEICE Transactions on Fundamentals E79-A(9), 1338–1353 (1996)
6. Tsai, J., Lo, N., Wu, T.: Secure Delegation-Based Authentication Protocol for Wireless Roaming Service. IEEE Communication Letters 16(7), 1100–1102 (2012)

An Improvement of Encrypted Remote User Authentication Scheme by Using Smart Card*

Jongho Mun[1], Jiye Kim[1], Woongryul Jeon[1], Youngsook Lee[2], and Dongho Won[1,**]

[1] School of Information and Communication Engineering, Sungkyunkwan University, Korea
{jhmoon,jykim,wrjeon,dhwon}@security.re.kr
[2] Department of Cyber Investigation Police, Howon University, Korea
ysooklee@howon.ac.kr

Abstract. Remote user authentication scheme is one of the most convenient authentication schemes to deal with secret data over insecure channels. In 2012, Yassin et al. proposed encrypted remote user authentication scheme by using smart card. They claimed that their scheme is secure against various attacks. In this paper, however, we display that their scheme is insecure and vulnerable to outsider attack, smart card stolen attack, offline password guessing attack, and masquerading attack. To overcome the drawback, we propose a new encrypted remote user authentication scheme by using smart card.

Keywords: smart card, remote user authentication, outsider attack, offline password guessing attack, masquerading attack.

1 Introduction

In 1981, Lamport [1] proposed a remote authentication scheme with insecure communication. Lamport's scheme resists a replay attack. However, it needed a password table for verifying the legitimacy of a login user. For this reason, smart card-based remote user authentication schemes are becoming day by day more popular. One of the benefits of the smart card-based authentication scheme is that a server does not have to keep a password table. This means that administrative overhead of server remarkably reduced. In the view of the fact that several remote user authentication schemes using smart card [2][3][4][5][6][7][8] have been proposed. In 2004, Das et al. [9] proposed a dynamic identity based remote user authentication scheme using smart cards which is secure against replay attack, password guessing attack, forgery attack, dictionary attack, and identity theft. However, their scheme is vulnerable to various attacks. In 2009, Wang et al. [10] presented a more secure dynamic ID-based remote user authentication scheme and

* This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2013-H0301-13-3007) supervised by the NIPA(National IT Industry Promotion Agency).

** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_64, © Springer-Verlag Berlin Heidelberg 2014

demonstrated the weaknesses of Das et al.'s scheme such as masquerade attack and lack mutual authentication. However, Wang et al.'s scheme suffers from malicious attacks and has some possible security risks.

Recently, Yassin et al. [11] proposed an improvement of Wang et al.'s scheme and demonstrated that Wang et al.'s scheme is still insecure and vulnerable to password guessing attack, DOS attack and server impersonate attack. However, in this paper, we find that Yassin et al.'s scheme is vulnerable to outsider attack, smart card stolen attack, off-line password guessing attack, and masquerading attack. We show that an outsider adversary can obtain the secret key of the server and we proposed more secure smart card-based remote user authentication scheme that overcomes these security vulnerabilities

This paper extended abstract of "Cryptanalysis of Encrypted Remote User Authentication Scheme by using Smart Card" published in conference proceeding of the CSA 2013 [12]. The rest of the paper is organized as follows. We begin by reviewing Yassin et al.'s encrypted remote user authentication scheme in section 2. Then in section 3, we describe security weaknesses in Yassin et al.'s scheme. Our proposed scheme is presented in section 4. Security analysis of our proposed scheme is given in section 5. Finally, we conclude this paper in Section 6.

2 Review in Yassin et al.'s Scheme

Yassin et al.'s scheme consists of four phases: registration, login, authentication, and password change phases.

2.1 Registration Phase

In this phase, the user U_i initially registers with the remote server S as follows:

1. User U_i sends his selected identity ID_i and hashed password $h(PW_i)$ to the remote server S over a secure channel.
2. After receiving $\{ID_i,\ h(PW_i)\}$, S computes $N_i = h(PW_i) \| h(X_S) \oplus h(ID_i)^{X_S}$, and $M_i = h(PW_i) \oplus h(X_S)$, where X_S is a secret key kept by S in private.
3. S stores the secure information $\{h(\cdot),\ N_i,\ M_i\}$ into a new smart card SC and sends a smart card to user U_i via a secure channel.

2.2 Login Phase

When a user U_i wants to login S, U_i inserts his smart card into the card reader and inputs his password PW_i. Then, smart card performs the following computations:

1. Smart card SC computes $h(X_S) = M_i \oplus h(PW_i)$ and $Z' = h(PW_i) \| h(X_S)$.
2. Upon computing $\{h(X_S),\ Z'\}$, SC generates a random number r_i and computes $K_i = h(r_i \oplus Z')$, $C_i = K_i \oplus (Z' \oplus N_i)^{r_i}$, and $f_i = h(ID_i)^{r_i}$.
3. Then, SC calculates $CID_i = Z' \oplus h(T \oplus r_i)$, where T is the current time stamp of the input device.

4. SC encrypts (r_i, T, N_i, CID_i) by using K_i.
5. After encrypting, SC sends login request message M = {C_i, f_i, $E_{K_i}(r_i$, T, N_i, $CID_i)$} to the remote server.

2.3 Authentication Phase

After receiving login request message at time T', S performs the following steps:

1. S computes $K_i = C_i \oplus f_i^{X_S}$, and decrypts $E_{K_i}(r_i, T, N_i, CID_i)$.
2. After decrypting, S checks the freshness of time stamp T. If T'- $T \leq \Delta T$ contains, where ΔT is the expected legal time interval for transmission delay, S persists the next step. Otherwise, S rejects the session.
3. S computes $Z'' = CID_i \oplus h(T \oplus r_i) = h(PW_i) \oplus h(X_S)$, and checks whether $(N_i \oplus Z'')^{r_i}$ is equal to $f_i^{X_S}$. If it is true, S accepts the user U_i's login request.
4. Then, S computes $a' = h(Z'' \| r_i \| T')$ and sends message $M' = E_{K_i}(a', T')$ to U_i.
5. When U_i receives the message $M' = E_{K_i}(a', T')$ at time T'', U_i checks whether T''- $T' \leq \Delta T$, where ΔT is the expected legal time interval for transmission delay. If T' is not fresh, U_i overthrows the message M' and terminates this phase. Otherwise U_i decrypts message M' by using K_i, and computes $a = h(Z' \| r_i \| T')$. Finally, U_i compares a with a'. If it is the case, U_i decides that the remote server S is authenticated.

2.4 Password Change Phase

When U_i wants to change his password from PW_i to PW_i^n, U_i implores this phase. The password change phase needs to pass the following steps:

1. User U_i must have executed the above login and authentication phase. The remote server S authenticates his old password PW_i.
2. After the successful mutual authentication, U_i enters his new password PW_i^n. Then, SC computes $N_i^n = N_i \oplus h(PW_i) \| h(X_S) \oplus h(PW_i^n) \| h(X_S)$ and replaces the old N_i with the new N_i^n.

3 Security Flaws in Yassin et al.'s Scheme

In this section, we point out security weaknesses of Yassin et al.'s scheme under following two assumptions.

1. An adversary can intercept all messages communicated among SC and S.
2. An adversary can steal smart card of legitimate user U, and he can obtain the parameter of U's SC.

Now, we show the security vulnerabilities of Yassin et al.'s scheme.

3.1 Outsider Attack

Any adversary U_a who is the legal user and owns a smart card, can get information $\{h(\cdot),\ N_a,\ M_a\}$, then he computes $h(X_S) = M_a \oplus h(PW_a)$ which is same for each legal user and is very sensitive information. Then, an adversary can compute secret key X_S of remote server S.

1. The adversary U_a calculates $h(X_S) = M_a \oplus h(PW_a)$, and then U_a obtains $h(ID_a)^{X_S} = N_a \oplus h(PW_a) \| h(X_S)$.
2. After obtaining $\{h(ID_a)^{X_S}\}$, U_a assumes that $R_0 = ID_a$, and i = 1.
3. Then, U_a calculates $R_i = h(R_{i-1})$. If $R_i = h(ID_a)^{X_S}$, then i = X_S else i = i + 1
4. Repeat 3.

3.2 Smart Card Stolen and Offline Password Guessing Attack

Smart card stolen attack means an adversary who possessed with smart card performs any operation which the smart card and obtains any information. If an outsider adversary U_a steals the SC of legitimate user U_i and obtains the parameters N_i and M_i, then he can easily computes out the hash value of the password of the real user U_i by computing $M_i \oplus h(X_S)$. Now, an adversary performs an off-line password guessing to get the current password of the user.

1. The adversary calculates $h(PW_i) = M_i \oplus h(X_S)$.
2. The adversary selects a random password $PW_i{}^*$, then calculates $h(PW_i{}^*)$ and compares it with $h(PW_i)$. If it is true, the adversary infers that $PW_i{}^*$ is user U_i's password. Otherwise, the adversary selects another password nominee and performs the same processes, until he locates the valid password.

3.3 Masquerading Attack

An outsider adversary U_a can easily masquerades as to remote server because he knows secret key X_S of remote server. If an outsider adversary intercepts login request message $\{C_i,\ f_i,\ E_{K_i}(r_i,\ \mathrm{T},\ N_i,\ CID_i)\}$ that the user U_i sends to the server S, he uses this knowledge of X_S and computes $K_i = C_i \oplus f_i{}^{X_S}$. Then an adversary decrypts $E_{K_i}(r_i,\ \mathrm{T},\ N_i,\ CID_i)$. Therefore, he can easily masquerade as to user U_i or server S.

4 Our Proposed Scheme

In this section, we propose a new encrypted remote user authentication scheme which improves on Yassin et al.'s scheme. The notations used in our proposed scheme are summarized as Table 1.

Table 1. Notations used in our proposed scheme

Notation	Description
U	A user
S	A remote server
ID, PW, SC	U's identity, password, and smart card
$\oplus$	The bitwise XOR operation
$\Vert$	String concatenation
X_S	S's secret key, which is kept secret and only known by S
$h(\cdot)$	A collision resistant one-way hash function
g	A generator
p	A large prime number
Q	A random value unique to the user

Our scheme consists of four phases as the same as in Yassin et al.'s scheme. However, before the registration phase, the server S decides on the following system parameters: a collision resistant one-way hash function h, a cryptographic key X_S, a large prime number p, and a generator g of $Z_p{}^*$.

4.1 Registration Phase

In this phase, the user U_i initially registers with the remote server S as follows:

1. User U_i chooses his identity ID_i and password PW_i, then computes $h(PW_i)$ and sends $\{ID_i, h(PW_i)\}$ to the remote server S over a secure communication channel.
2. Upon receiving $\{ID_i, h(PW_i)\}$, S computes $N_i = h(PW_i) \Vert h(X_S \Vert Q_i) \oplus g^{X_S * Q_i}$ mod p, $M_i = h(PW_i) \oplus h(X_S \Vert Q_i)$, where X_S is a secret key kept by S in private, and $L_i = Q_i \oplus h(PW_i)$ where Q_i is a random value unique to the user U_i.
3. S stores the secure information $\{h(\cdot), N_i, M_i, L_i, \text{g}, \text{p}\}$ into a new smart card SC and sends a smart card to user U_i via a secure channel.

4.2 Login Phase

When a user U_i wants to login S, U_i inserts his smart card into the card reader and enters his identity ID_i, and password PW_i. Then, smart card performs the following steps:

1. Smart card SC computes $h(X_S \Vert Q_i) = M_i \oplus h(PW_i)$, $Z' = h(PW_i) \Vert h(X_S \Vert Q_i)$, and $Q_i = L_i \oplus h(PW_i)$.
2. After computing $\{h(X_S \Vert Q_i), Z', Q_i\}$, SC generates a random number r_i and computes $K_i = h(r_i \oplus Z')$, $C_i = K_i \oplus (Z' \oplus N_i)^{r_i}$ mod p, and $f_i = g^{Q_i * r_i}$ mod p.
3. Then, SC calculates $CID_i = Z' \oplus h(T \oplus r_i)$, where T is the current time stamp of the input device.

4. SC encrypts (r_i, T, N_i, CID_i) by using K_i.
5. After encrypting, SC sends login request message M = {C_i, f_i, $E_{K_i}(r_i$, T, N_i, $CID_i)$} to the remote server S

4.3 Authentication Phase

After receiving login request message at time T', S performs the following steps:

1. S computes $K_i = C_i \oplus f_i^{X_S}$ mod p, and decrypts $E_{K_i}(r_i, T, N_i, CID_i)$.
2. After decrypting, S checks the freshness of time stamp T. If T'- $T \leq \Delta T$ contains, where ΔT is the maximum allowed time difference between T and T', then S persists the next step. Otherwise, S rejects the session.
3. S computes $Z'' = CID_i \oplus h(T \oplus r_i) = h(PW_i) \oplus h(X_S \| Q_i)$, and checks whether $(N_i \oplus Z'')^{r_i}$ mod p is equal to $f_i^{X_S}$ mod p. If it holds, S accepts the user U_i's login request.
4. Then, S computes $a' = h(Z'' \| r_i \| T')$ and sends message $M' = E_{K_i}(a', T')$ to U_i.
5. When U_i receives the message $M' = E_{K_i}(a', T')$ at time T'', U_i checks whether T''- $T' \leq \Delta T$, where ΔT is the maximum allowed time difference between T' and T'',. If T' is not fresh, U_i overthrows the message M' and terminates this session. Otherwise U_i decrypts message M' by using K_i, and computes $a = h(Z' \| r_i \| T')$. Finally, U_i compares a with a'. If it is true, U_i decides that the remote server S is authenticated.

4.4 Password Change Phase

Password change phase is same as Yassin et al.'s scheme. SC just performs the following step: After the successful mutual authentication, U_i enters his new password PW_i^n. Then, smart card SC computes $h(X_S \| Q_i) = M_i \oplus h(PW_i)$, $N_i^n = N_i \oplus h(PW_i) \| h(X_S \| Q_i) \oplus h(PW_i^n) \| h(X_S \| Q_i)$, $M_i^n = h(X_S \| Q_i) \oplus h(PW_i^n)$ and $L_i^n = L_i \oplus h(PW_i) \oplus h(PW_i^n)$ and replaces the old N_i, M_i, L_i with the new N_i^n, M_i^n, L_i^n.

5 Security Analysis of Our Proposed Scheme

In this section, we demonstrate that our scheme can withstand several possible attacks under two assumptions defined in section 3. The security of our proposed scheme is based on the hardness of discrete logarithm problem and the one-way hash function.

5.1 Outsider Attack

A legal but malicious adversary U_a can get $h(X_S \| Q_a)$, Q_a, and $g^{X_S * Q_a}$ mod p from his own smart card. However, U_a may not be able to compute X_S from $g^{X_S * Q_a}$ mod p, because it is very difficult which relies on the complexity of computing discrete logarithms over finite fields and Q_a is not the same for all users. Therefore, the proposed scheme is secure against outsider attack:

5.2 Smart Card Stolen Attack

An adversary can extract security parameters $\{h(\cdot), N_i, M_i, L_i$, g, p$\}$ from legitimate user U_i's smart card. However, this information does not help him. He cannot obtain the value of server S's secret X_S and user U_i's secret parameter Q_i. Thus, the proposed scheme is secure against smart card stolen attack.

5.3 Offline Password Guessing Attack

After obtaining the parameter on legitimate user U_i's smart card, an adversary can intercept to login request message $\{C_i, f_i, E_{K_i}(r_i$, T, $N_i, CID_i)\}$ during the login transaction, and try to guess out PW_i, Q_i, r_i, and X_S. However, an adversary cannot get useful information about the user U_i's password from these values, because other information is encrypted by U_i's secret key K_i. Therefore, the proposed scheme can resist offline password guessing attack.

5.4 Masquerading Attack

In our proposed scheme, an adversary U_a cannot play a role of the server S. He can know only $h(X_S \| Q_a)$, Q_a, and $g^{X_S * Q_a}$ mod p from his own smart card. Thus, he cannot play a role of server S by this values and he fails to obtain the values $\{X_S, K_i\}$. They are used to decrypt the ciphertext $E_{K_i}(r_i$, T, $N_i, CID_i)\}$ sent by U_i where K_i is calculated by $K_i = C_i \oplus {f_i}^{X_S}$ mod p. Thus, our proposed scheme is secure against server masquerading attack. When an adversary U_a wants to masquerade the valid user U_i, he needs to fake a legitimate login request message $\{C_i, f_i, E_{K_i}(r_i$, T, N_i, $CID_i)\}$, in which K_i = h$(r_i \oplus Z')$, $C_i = K_i \oplus (Z' \oplus N_i)^{r_i}$ mod p, and $f_i = g^{Q_i * r_i}$ mod p, $CID_i = Z' \oplus h(\mathrm{T} \oplus r_i)$. However, an adversary cannot get the server S's secret key X_S and the user U_i's unique value Q_i and fails to fake such a message.

6 Conclusion

In 2012, Yassin et al. proposed an encrypted remote user authentication scheme by using smart card and demonstrated that it's resistance to famous attacks. However, Yassin et al.'s scheme is vulnerable to outsider attack, smart card stolen attack, offline password guessing attack, and masquerade attack. In this paper, we propose an improved scheme to solve the vulnerabilities in Yassin et al.'s scheme. The security analysis explains that our improved scheme rectifies the weaknesses of Yassin et al.'s scheme.

References

1. Lamport, L.: Password authentication with insecure communication. Communications of the ACM, 770–772 (1981)
2. Chein, H.-Y., Chen, C.H.: A remote authentication scheme preserving user anonymity. Advanced Information Networking and Application 2, 245–248 (2005)
3. Lee, Y., Nam, J., Won, D.: Security enhancement of a remote user authentication scheme using smart cards. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4277, pp. 508–516. Springer, Heidelberg (2006)
4. Nam, J., Kim, S., Park, S., Won, D.: Security analysis of a nonce-based user authentication scheme using smart cards. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 90(1), 299–302 (2007)
5. Yi, W., Kim, S., Won, D.: Smart Card Based AKE Protocol Using Biometric Information in Pervasive Computing Environments. In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2009, Part II. LNCS, vol. 5593, pp. 182–190. Springer, Heidelberg (2009)
6. Jin, Q., Lee, K., Won, D.: Cryptanalysis of a two-factor user authentication scheme over insecure channels. In: ISA 2012 (2012)
7. He, D., Wu, S.: Security flaws in smart card based authentication scheme for multi server environment. Wireless Personal Communications, 0929–6212 (2012)
8. Son, K., Han, D., Won, D.: A Privacy-Protecting Authentication Scheme for Roaming Services with Smart Cards. IEICE Transactions on Communications 95(5), 1819–1821 (2012)
9. Das, M.L., Saxena, A., Gulati, V.P.: A dynamic ID-based remote user authentication scheme. IEEE Transactions on Consumer Electronics 50(2), 629–631 (2004)
10. Wang, Y.Y., Liu, J.Y., Xiao, F.X., Dan, J.: A more efficient and secure dynamic ID-based remote user authentication scheme. Computer Communications 4(32), 583–585 (2009)
11. Yassin, A.A., Jin, H., Ibrahim, A., Zou, D.: Encrypted Remote User Authentication Scheme by Using Smart Card. In: Wang, F.L., Lei, J., Gong, Z., Luo, X. (eds.) WISM 2012. LNCS, vol. 7529, pp. 314–323. Springer, Heidelberg (2012)
12. Mun, J., Kim, J., Jeon, W., Lee, Y., Won, D.: Cryptanalysis of Encrypted Remote User Authentication Scheme by Using Smart Card. In: Jeong, H.Y., Yen, N.Y., Park, J.J(J.H.) (eds.) Advanced in Computer Science and Its Applications. LNEE, vol. 279, pp. 423–428. Springer, Heidelberg (2014)

Countermeasure on Password-Based Authentication Scheme for Multi-server Environments*

Youngsook Lee[1], Jiye Kim[2], and Dongho Won[2,**]

[1] Department of Cyber Investigation Police, Howon University, Korea
ysooklee@howon.ac.kr
[2] School of Information and Communication Engineering, Sungkyunkwan University, Korea
{jykim,dhwon}@security.re.kr

Abstract. Recently, Tan proposed a remote user authentication scheme suite for multi-server environments, in which users can be authenticated using a single password shared with the registration center. A fundamental requirement for password-based authentication is security against off-line dictionary attacks. However, Tan's scheme fails to meet the requirement. In this paper, we report this security problem with Tan's scheme and show how to solve it.

Keywords: authentication scheme, smart card, password, off-line password guessing attack, multi-server environments.

1 Introduction

In 2011, Tan [8] proposed an efficient remote user authentication scheme suited for multi-server environments [1, 2, 3, 4, 9, 10, 11, 12, 13]. In its article, Tan claims that the user can be authenticated by all servers included in multi-server environments using a single password shared with the registration center and establishes the session key to be shared with between the server and the user.

In addition to making this claim, Tan claims to exhibit various merits with its scheme: (1) it allows the user to register only once with the registration center and then he/she is able to gain access to all servers included in multi-server environments without registering with every single server; (2) it does not require any server and the registration center to maintain a password table for verifying the legitimacy of login users; (3) it allows users to choose and change their passwords according to their liking and hence gives more user convenience; (4) it allows the user to change its password freely after assuring the legality of it; (5) it is extremely efficient in terms of the computational cost since the protocol participants perform only a few hash function operations; (6) it allows the user two factor security [8].

First of all, a fundamental requirement for password-based authentication is security against off-line password guessing attacks [16, 17]. However, Tan's scheme

* This work was supported by Howon University in 2014.
** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_65, © Springer-Verlag Berlin Heidelberg 2014

fails to meet the requirement. This paper is a revised and expanded version of [15]. In this paper, we report this security problem with Tan's scheme. What we do in this work is to report these security vulnerabilities of Tan's scheme.

The remainder of this paper is organized as follows. Section 2 reviews Tan's remote user authentication scheme. Section 3 presents our attacks on Tan's scheme and offers a security patch for the scheme.

2 Review of Tan's Password Authentication Scheme

This section reviews a remote user authentication scheme proposed by Tan [8]. The scheme participants include a registration center, a remote user, and multiple servers. For simplicity, we denote the registration center by *RC*, the remote user by U_i, and the servers by S_1, S_2, ..., S_n. The scheme assumes that the registration center *RC* is a trust party responsible for securely delivering the secret keys to be shared with between U_i and S_j . Tan's scheme consists of four phases: initialization phase, registration phase, login phase, and authentication phase. The initialization phase is processed when the server who wants to join to the system registers with the registration center. The registration phase is performed only once per user when a new user registers itself with the registration center. The login and the authentication phases are carried out whenever a user wants to gain access to each server included in multi-server environments. Before the registration phase is performed for the first time, the registration center *RC* decides on the following system parameters: a one-way hash function h and two cryptographic keys x and y. The keys x and y are shared securely with the registration center. The notation in Table 1. is employed throughout this paper and a more detailed description follows:

Table 1. Notation

PW_i	Password of an entity U_i
ID_i	Identity of an entity U_i
SID_j	Identity of an entity S_j
t_1, t_2, t_3, t_4	Timestamp
$E_K(X)$	Encryption of X using an asymmetric key K
$D_K(X)$	Decryption of X using an asymmetric key K
x, y	A cryptographic key
$h()$	One-way hash function
$\|\|$	Concatenation operation
$\oplus$	XOR operation
m, n, ξ	Random number

2.1 Initialization Phase

This is the phase where a new registration of a user takes place. The registration phase is invoked whenever a server wants to join this group. During this phase, the registration center *RC* and the server S_j perform the following running:

Step 1. A server S_j who wants to registration with the system submits its identity SID_j to the registration center RC via a secure channel.

Step 2. After receiving S_j's identity SID_j, RC computes ρ_j as $\rho_j = h(SID_j||y)$ and sends $< \rho_j >$ to RC through a secure channel.

2.2 Registration Phase

This is the phase where a new registration of a user takes place. The registration proceeds as follows:

Step 1. A user U_i, who wants to register with the registration center RC, submits a registration request, consisting of its identity ID_i, to the registration center RC via a secure channel.

Step 2. Upon receiving the request $<ID_i>$, RC computes $K_i = h\ (ID_i||x)$ and sends $<K_i>$ to user U_i.

Step 3. Now the user U_i chooses its password PW_i at will and computes $B_i = K_i \oplus h\ (ID_i||PW_i)$. Then U_i stores the values $<B_i, ID_i, h\ ()>$ on its smart card.

2.3 Login Phase

Step 1. When U_i wants to log in to the system, he inserts his smart card into a card reader and enters his identity ID_i and password PW_i.

Step 2. Given ID_i and PW_i, the smart card generates the random number m and computes

$$T_i = B_i \oplus h\ (ID_i||PW_i),$$
$$\pi = g^m,$$
$$C_1 = h\ (T_i \oplus SID_j) \oplus \pi,$$
$$C_2 = h\ (C_1||\pi).$$

The smart card then sends the login request message $<ID_i, SID_j, C_1, C_2>$ to the server S_j .

Step 3. When the login request arrives $<ID_i, SID_j, C_1, C_2>$, the server S_j first chooses the random number ξ and computes $V_1 = h\ (\rho_j||ID_i) \oplus \xi$ and $V_2 = h\ (C_1||C_2||V_1||\xi)$. Then S_j sends $<ID_i, SID_j, C_1, C_2, V_1, V_2>$ to the registration center RC.

2.4 Authentication Phase: Server and Registration Phase

With the login request message $<ID_i, SID_j, C_1, C_2, V_1, V_2>$, the scheme enters the authentication phase during which S_j and RC perform the following steps:

Step 1. After receiving <ID_i, SID_j, C_1, C_2, V_1, V_2> from S_j , the registration center computes

$$\pi' = h\,(h\,(ID_i||x)\,||SID_j) \oplus C_1,$$
$$\xi' = h\,(h\,(SID_j\,||y)\,||ID_i) \oplus V_1,$$
$$\tau = \pi'^{\,h(\pi')}.$$

The registration center RC verifies that: (1) C_2 equals $h\,(C_1||\pi')$ and (2) V_2 equals $h\,(C_1||C_2||V_1||\xi')$. If both of these conditions hold, RC believes that the responding party is the genuine user and server, respectively. Otherwise, RC aborts its login attempt.

Step 2. After that, RC computes the response

$$Z_1 = h\,(h\,(SID_j\,||T_i)\,||ID_i) \oplus \pi' \oplus \xi',$$
$$Z_2 = h\,(Z_1 \oplus \xi'||\,h\,(ID_i\,||h\,(SID_j||y)\,||\tau\,) \oplus \xi',$$
$$Z_3 = h\,(Z_2||\xi'||h\,(ID_i||h\,(SID_j||y)))\oplus \tau.$$

Then RC sends <Z_1, Z_2, Z_3>to the server S_j .

Step 3. Having received <Z_1, Z_2, Z_3>from RC, S_j computes $\tau' = h\,(Z_2||\xi||\,h\,(ID_i||\rho_j))\oplus Z_3$. S_j verifies that Z_2 equals $h\,(Z_1\oplus \xi||h\,(ID_i||h\,(SID_j||y)\,||\,\tau'\,) \oplus \xi$. If they are not equal, S_j believes that he is talking to illegal registration center and aborts the scheme. Otherwise, S_j continues the following procedures.

2.5 Authentication Phase: Server and User

Step 1. Upon believing that RC is authentic, S_j chooses the random number n and computes

$$\mu = g^n$$
$$V_3 = h\,(ID_i||\,\rho_j) \oplus h\,((Z_1 \oplus \xi)\,||SID_j),$$
$$V_4 = h\,((Z_1 \oplus \xi)\,||ID_i||V_3||\mu),$$
$$V_5 = h\,((Z_1 \oplus \xi)\,||ID_i||SID_j) \oplus \mu,$$
$$SK = h\,((Z_1 \oplus \xi)\,||\,\tau^{\,n}||h\,(ID_i||\,\rho_j).$$

Then the server S_j sends <V_3, V_4, V_5> to the user U_i.

Step 2. After receiving <V_3, V_4, V_5>, U_i computes

$$\mu' = V_5 \oplus h\,((h\,(SID_j\,||T_i||ID_i) \oplus \pi)||ID_i||SID_j),$$
$$SK = h\,(h\,(SID_j||T_i||ID_i \oplus \pi\,||\mu'^{\,mh(\pi)}||V_3) \oplus h\,(h\,(SID_j\,||T_i||ID_i) \oplus \pi||\,SID_j).$$

Then user U_i verifies that V_4 equals $h\,((h\,(SID_j\,||T_i||ID_i) \oplus \pi)||ID_i||V_3||\,\mu')$. If they are equal, U_i believes S_j as authentic. Otherwise, U_i aborts its login attempt.

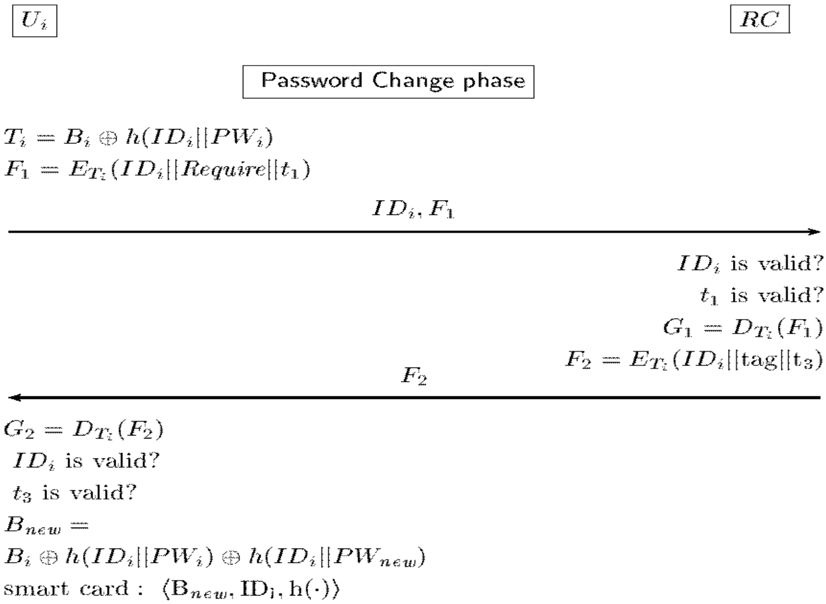


Fig. 1. The password change phase on Tan's scheme

2.6 Password Change Phase

If U_i wants to change its password, U_i inserts its smart card and its identity ID_i and password PW_i. The user U_i issues a *Require* of replacing old password with a new password. A high level depiction of the scheme is given in Fig. 1. The smart card executes the following steps.

Step 1. Given ID_i, *Require*, and PW_i, the smart card generates a timestamp t_1 computes $T_i = B_i \oplus h(ID_i||PW_i)$ and $F_1 = E_{T_i}(ID_i||Require||t_1)$.

Then U_i sends the password request message $<ID_i, F_1>$ to the registration center *RC*.

Step 2. After receiving the message $<ID_i, F_1>$, *RC* first acquires the current timestamp t_2 and computes G_1 as $G_1 = D_{Ti}(F_1)$. Then *RC* verifies that: (1) ID_i is valid, (2) $t_2 - t_1 \leq \Delta t$ where Δt is the maximum allowed time interval for transmission delay. If both of these conditions hold, *RC* believes that the responding party is the genuine user and makes a tag which denotes Yes as the response for the request. Otherwise, *RC* makes a tag which denotes No. Now *RC* generates a new timestamp t_3, computes $F_2 = E_{Ti}(ID_i||\mathrm{tag}||t_3)$ and sends to the response message $<F_2>$ to the user U_i.

Step 3. Having received $<F_2>$ from *RC*, U_i generates the current timestamp t_4 and computes G_2 as $G_2 = D_{T_i}(F_2)$. Then U_i verifies that: (1) ID_i is valid, (2) $t_4 - t_3 \leq \Delta t$ where Δt is the maximum allowed time interval for transmission delay (3) Tag is yes?. If any of these is untrue, U_i stop the password change. Otherwise, the user U_i

chooses a new password PW_{new}, computes $B_{new} = B_i \oplus h\ (ID_i||PW_i) \oplus h\ (ID_i||PW_{\text{new}})$. Finally, U_i replaces B_i with B_{new} on the smart card.

3 Cryptanalysis of Tan's Scheme

Here we point out a security problem is that Tan's scheme has not password security. We interpret this problem as the vulnerability of the scheme to off-line password guessing attack. A few years ago, we demonstrate the password security of remote user authentication schemes using smart cards [14]. To analyze the security of remote user authentication schemes using smart cards, we need to consider the capabilities of the attacker. First, we assume that the attacker has complete control of every aspect of all communications between the server and the remote user. That is, he/she may read, modify, insert, delete, replay and delay any messages in the communication channel. Second, he/she may try to steal a user's smart card and extract the information in the smart card by monitoring the power consumption of the smart card [5, 6]. Third, he/she may try to find out a user's password. Clearly, if both (1) the user's smart card was stolen and (2) the user's password was exposed, then there is no way to prevent the attacker from impersonating the user. However, a remote user authentication scheme should be secure if only one of (1) and (2) is the case. So the best we can do is to guarantee the security of the scheme when either the user's smart card or its password is stolen, but not both. This security property is called two-factor security [9]. In this section we point out that Tan's scheme does not achieve its main security against off-line password guessing attack [14].

3.1 Off-Line Password Guessing Attack

Tan [8] claims that its authentication scheme prevents an attacker from learning some registered user's password via an off-line password guessing attack. But, unlike the claim, Tan's scheme is vulnerable to an off-line password guessing attack mounted by extracting the secret information from a smart card [9]. In Tan's scheme, assume that an attacker has stolen the U_i's smart card or gained access to it and extracted the secret value(B_i) stored in it by monitoring its power consumption [5, 6]. Now the attacker U_a obtains the value B_i stored in U_i's smart card. Then the following description represents our off-line guessing attack mounted by the attacker U_a against U_i's password: The attacker U_a, who wants to find out PW_i, now guesses possible passwords and checks them for correctness.

Step 1. As preliminary step, the attacker U_a, who has obtained B_i stored in its smart card.

Step 2. As usual, the password change phase begins when the user U_i sends the password change request message $<ID_i, F_1>$to the registration center.

Step 3. After accepting the registration center as authentic by verifying the authenticity of the received message $<ID_i, F_1>$ from U_i, RC sends $<F_2>$ as a password change response to the user U_i.

Step 4. However, at that moment, U_a reads these exchanged messages between the registration center and the remote user and thereby obtains all of them. An attacker U_a, who has extracted B_i from U_i's smart card and obtained the values ID_i, F_1, and F_2 from the U_i's password change request message and RC's response message, can find out PW_i by employing the off-line password guessing attack.

Step 5. Now U_a makes a guess PW'_i for the password PW_i and computes

$$T'_i = B_i \oplus h\,(ID_i || PW_i) \oplus h\,(ID_i || PW'_i),$$
$$G'_1 = D_{T'_i}\,(F_1),$$
$$G'_2 = D_{T'_i}\,(F_2).$$

Then U_a verifies the correctness of PW_i by checking that ID_i is valid?. Notice that if PW'_i and PW_i are equal, and then ID_i ought to be identify.

Step 6. U_a repeats Step 5. using another guessed password until a correct password is found.

3.2 Preventing the Attack

We now figure out what is wrong with the scheme and how to fix it. The fixed scheme is given mainly to provide a better insight into the failure of Tan's scheme.

Flaws in the Scheme. The vulnerability of Tan's scheme to the password guessing attack is due to the following fact: to find out the password of the user, they suffice to obtain the information stored in its smart card and read the exchanged message between the registration center and the remote user. More concretely, the problem with Tan et al.'s scheme is that whoever obtains these values of B_i stored in U_i's smart card, the registration center's response F_2, and the part of the user U_i's login message F_1 can break the user U_i's password PW_i.

Countermeasure. The simple way to resolve the security problem with Tan's scheme would be to change the computations of the encryption key T_i to:

$$\alpha_i = h\,(T_i || \tau).$$
$$\tau = \pi^{h\,(\pi)}.$$

Our proposed scheme effectively defeats these kinds of attacks mentioned above. Even if the attacker obtains the information (i.e., B_i) stored in the smart card and the exchanged message between the server and the user, he/she can no longer find out the password of the user U_i. In the patched scheme, the only information related to passwords is α_i ($= h\,(T_i || \pi^{h\,(\pi)}$), but because $\pi^{h\,(\pi)}$ is the secret information that the user and the registration center only know, this value does not help the attacker to verify directly the correctness of guessed passwords. Thus, off-line password guessing attack would be unsuccessful against the proposed protocol.

Acknowledgements. This work was supported by Howon University in 2014. The primary (first) author is Youngsook Lee (ysooklee@howon.ac.kr).

References

1. Chang, C., Kuo, J.: An efficient multi-server password authenticated keysagreement scheme using smart cards with access control. In: IEEE Proceeding of the 19th International Conference on Advanced InformationNetworking and Applications, vol. 2, pp. 257–260 (2005)
2. Ku, W., Chang, S., Chiang, M.: Weaknesses of a remote userauthentication scheme using smart cards for multi-serverarchitecture. IEICE Transactions on Communications E88-B(8), 3451–3454 (2005)
3. Li, L., Lin, L., Hwang, M.: A remote passwordauthentication scheme for multi-server architecture using neuralnetworks. IEEE Transaction on Neural Networks 12(6), 1498–1504 (2001)
4. Lin, I., Hwang, M., Li, L.: A new remote userauthentication scheme for multi-server architecture. Future Generation Computer System 19, 13–22 (2003)
5. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
6. Messergers, T., Dabbish, E., Sloan, R.: Examining smart card securityunder the threat of power analysis attacks. IEEE Trans. Comput. 51(5), 541–552 (2002)
7. Sun, H.-M.: An effcient remote user authenticationscheme using smart cards. IEEE Transaction on Consumer Electronics 46(4), 958–961 (2000)
8. Tan, Z.: Improvement on A Password Authentication Scheme for Multi-server Environments. Journal of Convergence Information Technology 6(1), 218–228 (2011)
9. Tian, X., Zhu, R., Wong, D.: Improved efficient remote user authenticationschemes. International Journal of Network Security 4(2), 149–154 (2007)
10. Tsai, J.: Efficient multi-server authentication scheme based onone-way hash function without verification table. Computers & Security 27, 115–121 (2008)
11. Tsuar, W.: An enhanced user authentication scheme for multi-serverinternet services. Applied Mathematics and Computation 170, 258–266 (2005)
12. Tsaur, W.-J.: A flexible user authentication scheme for multi-server internet services. In: Lorenz, P. (ed.) ICN 2001. LNCS, vol. 2093, pp. 174–183. Springer, Heidelberg (2001)
13. Tsuar, W., Wu, C., Lee, W.: A smart card-based remotescheme for password authentication in multi-server Internetservices. Computer Standards & Interfaces 27, 39–51 (2004)
14. Lee, Y., Won, D.: Security vulnerabilities of a Remote userauthentication scheme using smart cards suited for a multi-server environment. In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2009, Part II. LNCS, vol. 5593, pp. 164–172. Springer, Heidelberg (2009)
15. Lee, Y., Kim, J., Won, D.: Security analysis onpassword-based authentication scheme for multi-server environments. In: The Second International Conferenceon Computers, Networks, Systems, and Industrial Applications, CNSI, pp. 97–101 (July 2012)
16. Jeon, W., Kim, J., Nam, J., Lee, Y., Won, D.: An Enhanced Secure Authentication Scheme with Anonymity for wireless Environments. IEICE Transaction on Communications E95-B(7), 2505–2508 (2012)
17. Yang, W.-H., Shieh, S.-P.: Password authenticationschemes withsmart card. Computers & Security 18(8), 727–733 (1999)

The Organization and Presentation of Video Continuity Using 3D Floating Technique in STEAM Education*

Jeongyeun Kim and Namje Park**

Major in Computer Education, Faculty of Science Education,
Graduate School, Jeju National University,
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea
namjepark@jejunu.ac.kr, inarasam@naver.com

Abstract. The purpose of this paper is to provide students on natural sciences track and science-focused track who are talented in science and mathematics and have interest in the convergence and application of related knowledge with the opportunity to interpret the rapidly changing media industry in terms of science and actually apply related technologies to grow into creative intellectuals of the future society. Furthermore, this course intends to encourage students to collect information for the coherent theme, the high-tech movies, and reinterpret them in terms of science to turn the media that had been felt distant into the subject of study and also produce movies of their interest to realize that the industry, technology, art, and science are indeed interrelated with each other.

Keywords: 3D, 3D Floating, Video Continuity, Elementary School, STEAM Education.

1 Introduction

When producing a high-tech movie using advanced scientific principles, it can become a movie that no one wants to see if the producer only depends on scientific knowledge. Also, subjects included in the curriculum but seem unrelated with each other such as writing, art, and music are actually quite inevitably interrelated with each other. If even a few elements of these subjects can be incorporated into a task, it will be possible to produce the outcome that everyone wants to see, realizing creative design and fulfilling the emotional touch of students that carried out the task. The educational goal of this course in relation to the curriculum is to enhance positive motivation for achievement for students concerning social and scientific tasks of their interest they would perform in the future by giving them the chance to have such an exciting and composite experience of a success.

---

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A4A01 013587).

** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_66, © Springer-Verlag Berlin Heidelberg 2014

The purpose of this paper is to provide students on natural sciences track and science-focused track who are talented in science and mathematics and have interest in the convergence and application of related knowledge with the opportunity to interpret the rapidly changing media industry in terms of science and actually apply related technologies to grow into creative intellectuals of the future society. Furthermore, this course intends to encourage students to collect information for the coherent theme, the high-tech movies, and reinterpret them in terms of science to turn the media that had been felt distant into the subject of study and also produce movies of their interest to realize that the industry, technology, art, and science are indeed interrelated with each other.

2 Analysis of STEAM Elements

2.1 Objectives

A. Content Objectives

1) Find and understand scientific principle used in complicated high-tech movies.
2) Devise and apply the method for producing high-tech movies inside the classroom using scientific principles.
3) Produce an original movie by combining knowledge from various fields.

B. Course Objectives (Suggest content objectives related with emotional touch)

1) Find scientific principles in various latest movies and reanalyze them logically.
2) Produce imaging equipment based on scientific principles.
3) Produce an original and new movie based on scientific principles.

2.2 Subject Elements of STEAM

A. S: Investigation of scientific principles used in high-tech movies and application of such principles in producing an imaging device
B. T: Investigation of the latest movie techniques design and production of related devices
C. E: Design and production of imaging device using scientific principles
D. A: Production of continuity for each team's movie and production of movies
E. M: Drawing of geometric figures in the device design and the analysis of parabolic equation in parabolic lamp

2.3 Phase Elements of STEAM

Suggest elements based on the frame of STEAM such as Co, Cd, and ET

A. Context (Co): Show the examples of recent movies made with advanced technology, show the example of the design of high-tech imaging equipment, suggest continuity for a movie that can be realized, show a variety of examples relating to the subject of study

B. Creative Design (Cd): Investigation of scientific principles and design of movie equipment using the principles, production of original continuity and movie and discussion
C. Emotional Touch (ET): Derive scientific principles and design a device applying the principles, production of creative movie based on continuity in each team and discussion about the applications

3 Understanding 3D Floating Technique and Organization and Presentation of Video Continuity

3.1 Truth of Hologram in the Movie (Movie Using 3D Floating Technique)

We can draw a rough picture in our minds about the technology in near future while watching SF (Science Fiction) movies. The director of the movie expresses his or her imagination through the movie, and such imagination stimulates the technology in the related area that we often see products or technologies that actualized the imagination in everyday living. You must remember the scene from “Iron Man”, the movie by Jon Favreau that enjoyed world-wide popularity where the hero uses hologram to design a special suite and test it directly in the holographic environment. The holographic technology that looked so cool in the movie... is it possible to build that sort of hologram with today's technology? Unfortunately, it is not. Well, then let's see how advanced today's holographic technology is and the essence of the holographic technology we see. To understand the technological meaning of the hologram, we need to take a look at where the word and hologram came from. According to the dictionary, "hologram" is a compound word made from "holos" which means "whole" and "gram" which means a picture. In other words, it means a "whole picture." Let's go back to the scenes from the movie in a more realistic view this time. I said it is difficult to realize the scene where the hero puts the holographic image in the air and directly tests the suit with it, but what if we do it another way? The answer is yes. To form an image in the air without any medium, it will take hundreds or thousands of laser light source and a supercomputer to send and process data that it is impossible in reality, but if we use the reflection technique that uses a medium that acts as a screen that is hardly seen with eyes or 3D floating technology (hologram-like technology) that uses direct projection, we can put an image in the air and move it as we wish like the hero in the movie.

Let's take a closer look at the hologram that uses principle of reflection. The hologram that uses the principle of reflection originated in approximately the 19th century. Magicians at the time put a ghost on their stages, and they used the optical illusion called "Pepper’s Ghost" to delude the audience.

The basic principle is that a transparent screen tilted to 45° is installed and an ordinary screen is installed underneath to reflect the image, and when the image is reflected on the transparent screen, it looked as though a person or an image is floating in the air. This technique which started from a very simple principle is evolving into a new type of holographic technique by combining various information technologies.

The hologram shown in “Iron Man” can be realized even by us today by combining various information technologies capable of interacting with the hologram. For example, the reflective hologram combined with various information technologies such as the motion sensor for controlling motion, the sound sensor for controlling sound, and tele-presence which transmits images in real time to enable conversation or conference using holograms can enable us today to put an image in the air to control it and use it to talk to a person at a distance in real time as if we are talking face to face as they do in the movies.

3.2 Imaging Device to Use for 3D Floating Technique in This Class

A transparent acrylic panel (1.2m×2.4m) was installed at 1m above the floor at 45°, and 9 sheets of whole paper were laid on the floor in front of the panel as the reflective screen, and the single-focus beam projector was used as the light source.

Fig. 1. Imaging device to use for 3D floating technique

3.3 Explaining the Example of the Continuity for Using the Advantage of 3D-Floating Hologram Technique

A transparent acrylic panel (1.2m×2.4m) was installed at 1m above the floor at 45°, and 9 sheets.

3.4 Advantages and Disadvantages of 3D Floating Technique and Application Method

The transparent Holo Screen (plastic foil) processed with a special technology is a completely transparent screen unlike the ordinary screen, and it can display a clean and 3-dimensional image by projecting the hologram image from front and back. The screen is transparent that the image displayed on it looks as though it is incorporated into the reality but it cannot be considered as complete 3-dimensional representation, and it also has the limit that it cannot be viewed from the side. Also, the resolution is high that the image is clean, however, is weak and dim. It is the type of 3-dimensional presentation shown in movies such as the A.I, the holographic expression mainly

Title: Versatile A Team: A, 5 Members

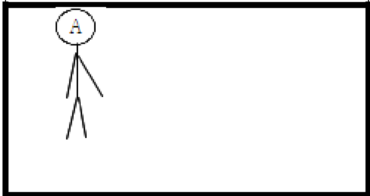
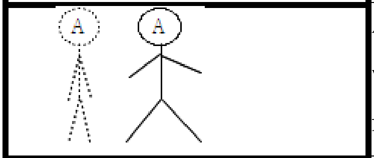
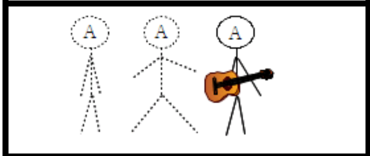
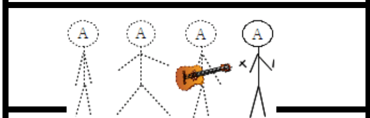
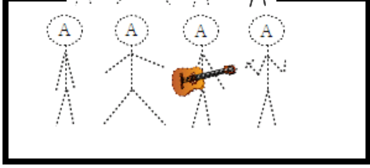
Cut	Video	Context	Audio	Time
1		A sings to the music and bows.	Record A's singing part only.	5 Minutes
2		A dances and bows while the hologram from Cut 1 is floating.	No sound	5 Minutes
3		A plays guitar and bows while the hologram from Cut 2 is floating.	Record guitar sound.	5 Minutes
4		A conducts and bows while the hologram from Cut 3 is floating.	No sound	5 Minutes
5		Float the image from Cut 4 to retain the same image quality.	No sound	5 Minutes

Fig. 2. Example of the continuity for using the advantage of 3D-floating hologram technique

shown inside the movie. The picture on the left below is the advertisement on the shop window, and the one on the right below gives the impression that the holographic image is actually talking to the viewer.

There is also the floating movie technique which uses a projector to project a 3D video on a special screen to generate a floating image, a visually 3-dimensional image. This is mainly used for shop window display and is effective for exhibition and showing. However, it requires sufficient space for it uses a separate projector. It can sufficiently satisfy the desired advertisement effect for it can easily capture the eyes of the people when it is applied to the window of the street shop as in the following picture. It can make viewers think as if a real person is changing clothes right in front of their eyes or simply show a football game. It added fun element that can holography can give when showing a football game by making the referee look as though he is popping out from the ground. This also has small limitation with colors but has the disadvantage that the whole image is dim.

4 Conclusion

The purpose of this paper is to provide students on natural sciences track and science-focused track who are talented in science and mathematics and have interest in the convergence and application of related knowledge with the opportunity to interpret the rapidly changing media industry in terms of science and actually apply related technologies to grow into creative intellectuals of the future society. Furthermore, this course intends to encourage students to collect information for the coherent theme, the high-tech movies, and reinterpret them in terms of science to turn the media that had been felt distant into the subject of study and also produce movies of their interest to realize that the industry, technology, art, and science are indeed interrelated with each other.

References

1. Ko, Y., An, J., Park, N.: Development of Computer, Math, Art Convergence Education Lesson Plans Based on Smart Grid Technology. In: Kim, T.-H., Stoica, A., Fang, W.-C., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H. (eds.) SecTech/CA/CES^3 2012. CCIS, vol. 339, pp. 109–114. Springer, Heidelberg (2012)
2. Kim, Y., Park, N.: The Effect of STEAM Education on Elementary School Student's Creativity Improvement. In: Kim, T.-H., Stoica, A., Fang, W.-C., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H. (eds.) SecTech/CA/CES^3 2012. CCIS, vol. 339, pp. 115–121. Springer, Heidelberg (2012)
3. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
4. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
5. Park, N.: Secure UHF/HF Dual-Band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
6. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. International Journal of Ad Hoc and Ubiquitous Computing 8(4), 205–219 (2011)
7. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
8. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. The Journal of Korea Information and Communications Society 35(5), 785–793 (2010)

STEAM Education Program: Small Organisms Creating Energy*

Soonwoo Park and Namje Park**

Department of Computer Education, Teachers College, Jeju National University, 61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea
{park0820,namjepark}@jejunu.ac.kr

Abstract. This paper examined various study case about piezo-electric elements and STEAM education. We develop This STEAM education using piezo-electric elements for middle school students. We make this program and teaching materials like electric circuit using piezo-electric elements by considering about student's intellectual level and interests. We offered this program to 12 middle school students in jeju island step by step for 10 period to evaluate the effectiveness of the developed program and teaching materials. Future work will include the follow-up and extended study. It will be expected that students will be able to develop their abilities by this STEAM education using new technology. The purpose of this paper is as follows; first, it is to devise a STEAM program for making an electronic product using the energy produced from a piezo-electric element targeting middle school students. Second, it is to develop a teaching and learning material that can be utilized in the educational program.

Keywords: STEAM, Piezo-electric Element, Elementary School, Teaching Method.

1 Introduction

The details of the activities to be carried out to fulfill the goal of this paper are as follows; first, make students understand piezo-electric element and produce an electronic product using the energy generated from piezo-electric element. Second, provide students with the chance to have actual experience through the STEAM materials for producing an electronic product using piezo-electric element. Third, develop the course guideline, power point activity log, and learning activity log after producing electronic products. Fourth, we applied STEAM materials in the actual class to increase satisfaction about the class. We investigated the subject using newspaper article, academic books, and academic papers and also analyzed curriculums and textbooks revised in 2009 in related subjects in order to determine the learning capability of students by age.

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A4A01 013587).

** Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 308,
DOI: 10.1007/978-3-642-54900-7_67, © Springer-Verlag Berlin Heidelberg 2014

We held a number of seminars among us and a consultative meeting with outside experts to review the feasibility of the contents and the applicability of the education program development plan we developed. Furthermore, the teaching materials to be applied to the education program were produced to reflect convergence in contents domain pursued by STEAM education and learner's characteristics in terms of educational psychology.

Major results of this study are as follows; first, the educational program is composed of 10 classes in total, and the program is mainly constituted with the following themes. ① Press to generate electricity. (Piezo-electric ceramic, energy harvesting, advantage of piezo-electric energy), ② Handling piezo-electric element. (Testing piezo-electric element, operating LED with piezo-electric element), ③ Drawing my character (Giving design advice, making engineering design), ④ Making my own piezo-electric element energy device. (Understanding energy block, producing piezo-electric element presser plate, making energy hand), ⑤ Making communication device using piezo-electric element (introduction to electronic parts (LED, buzzer) and design of electric circuit diagram, making my own communication device, sending information using communication device), ⑥ Finding the application of piezo-electric element in actual living (the actual application of piezo-electric element in actual living, developing and designing a new idea to be applied to actual living). Second, for the teaching and learning materials to be applied to the program, students are encouraged to understand the characteristics of piezo-electric element, design a device that directly generates energy based on the knowledge of the element, and devise and produce a model product to apply this energy in real life (When a person claps with the device equipped with piezo-electric element or two people mutually apply pressure to this device to generate electric energy, this energy is converted into the light energy through LED). This product has the characteristic of a device which generates and converts environment-friendly energy with which the participant can produce and confirm the energy by himself or herself. Third, teaching-learning course guide for teacher which will provide the guideline for teaching and learning while 10 classes are operated was developed along with the instruction guide for teacher and textbooks for students.

This educational program is expected to set and discover educational direction and potential in educational settings in relation to piezo-electric element and also make a great contribution to the development of high quality STEAM programs using piezo-electric element.

2 Background and Purpose of Program Development

2.1 Background for Program Development

The piezo-electric technology for resolving energy resource issue produces electricity by collecting energy wasted around us that it is in the lime light as the new type of energy source.

2.2 Purpose for Program Development

Recently, many researchers are studying the piezo-electric power generation technology which uses piezo-electric effect to produce electricity. Under the circumstances we developed the experience-based program that involves piezo-electric element for school curriculum.

2.3 Purpose of the Program

The purpose of this program is to have middle school students recognize the relationship between subjects using piezo-electric element, develop convergent (STEAM) thinking through experiential activities, and obtain outcomes at the end.

2.4 Phase Elements of STEAM Program

Suggest elements based on the frame of STEAM such as Co, Cd, and ET

A. Context (Co): Have you had the shoes that sparkled with light when you walk? The shoes do not sparkle when the battery dies. Do you know how to make them sparkle forever without a battery?
B. Creative Design (Cd): Making my own hand energy device and information and communication device
C. Emotional Touch (ET): Understanding the principle of piezo-electric element using hand energy device, communicating with the best friend with my own communication device

3 Process of Developing the Teaching Aid to be Applied to STEAM Program and the Completed Model (Example)

3.1 Electric Energy Generating Model Using Piezo-electric Element

When a person claps with the device equipped with piezo-electric element or two people mutually apply pressure to this device to generate electric energy, this energy is converted into the light energy through LED to generate the environment-friendly energy that can be visually manifested.

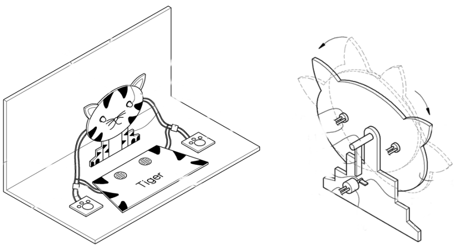


Fig. 1. Teaching Aid to be applied to STEAM Program and the Completed Model

3.2 Process of Developing the Teaching Aid To Be Applied to STEAM Program

Order	Process Name	Process Details	Tools	Required Time
1	Preparations	•Set up the working plan. •Prepare materials and tools required, •Wear comfortable clothes for working. •Prepare design drawing.		0.3
2	Cutting	•Draw lines on the material according to the design. •Cut the material to the required size.	Pencil, Knife, Ruler	0.5
3	Processing	•Solder. •Connect wires. •Make a hole.	Soldering Iron, Lead	0.5
4	Assembly	•Connect piezo-electric element device and character. •Mount presser plate and character on the support fixture.	Glue Gun	0.5
5	Evaluation	•Inspect and evaluate according to assembly diagram.		0.2

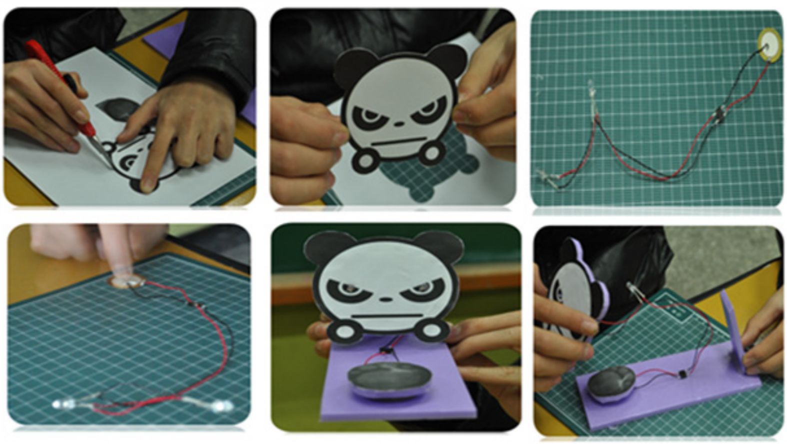


Fig. 2. Process of Developing the Teaching Aid to be applied to STEAM Program

① Character Design
Draw a design diagram by devising a character for a hand energy device.

② Cutting the design
- Using a knife, cut out the character design.
- Using a knife, cut out the presser plate design.

③ Development of a character
- Attach character design to the form board using glue gun.
- Using a knife, cut the form board along the character design.
- Attach presser plate design to the form board, and cut.
- Check the distance between the hand energy device character and the presser plate and make a support fixture with a form board to connect the character and plate.

④ Composition of circuit diagram
- Connect one of 2 wires to piezo-electric element and the other, bridge diode.
- Connect the other end of the bridge diode to LED.
- Connect two LEDs in parallel. After composing the electric circuit diagram for the hand energy device, check whether LED turns on by pressing and releasing the piezo-electric element.

⑤ Assembly
- Connect hand energy electric device to character face.
- Devise and make an effective presser tool for stable piezo-electric effect.
- Place the presser plate device to the support fixture plate using glue gun.
- Place character face device on the support fixture plate using glue gun. When the model is completed, check whether it is made as designed.

⑥ Evaluation
Let's develop and use hand energy device. Share what you felt about it through a presentation. Also, try to evaluate your own hand energy device.

4 Conclusion

The STEAM Education Program on the theme of bioenergy suggested in this study can increase students' interest in STEAM Education and bioenergy to resolve problems in real life and develop knowledge of and positive attitude toward science and technology. Furthermore, the program deals with the contents students have already learned during in the school curriculum, and therefore students can incorporate the contents they learned into the process of solving problems. Students can also develop self-directed learning capabilities and problem solving capabilities while planning, implementing, and evaluating the process of producing microbial fuel cell and designing biohouse by themselves.

References

1. Park, N., Ko, Y.: Computer Education’s Teaching-Learning Methods Using Educational Programming Language Based on STEAM Education. In: Park, J.J., Zomaya, A., Yeo, S.-S., Sahni, S., et al. (eds.) NPC 2012. LNCS, vol. 7513, pp. 320–327. Springer, Heidelberg (2012)

2. Ko, Y., An, J., Park, N.: Development of Computer, Math, Art Convergence Education Lesson Plans Based on Smart Grid Technology. In: Kim, T.-H., Stoica, A., Fang, W.-C., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H., et al. (eds.) SecTech/CA/CES^3 2012. CCIS, vol. 339, pp. 109–114. Springer, Heidelberg (2012)
3. Ko, Y., Park, N.: Experiment and Verification of Teaching Fractal Geometry Concepts Using a Logo-Based Framework for Elementary School Children. In: Kim, T.-H., Adeli, H., Slezak, D., Sandnes, F.E., Song, X., Chung, K.-I., Arnett, K.P., et al. (eds.) FGIT 2011. LNCS, vol. 7105, pp. 257–267. Springer, Heidelberg (2011)
4. An, J., Park, N.: The Effect of EPL Programming Based on CPS Model for Enhancing Elementary School Students' Creativity. In: Park, J.J(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. LNEE, vol. 181, pp. 237–244. Springer, Heidelberg (2012)
5. An, J., Park, N.: Computer Application in Elementary Education Bases on Fractal Geometry Theory Using LOGO Programming. In: Park, J.J., Arabnia, H., Chang, H., et al. (eds.) IT Convergence and Services. LNEE, vol. 107, pp. 241–249. Springer, Heidelberg (2011)
6. Park, N., Kwak, J., Kim, S., Won, D.H., Kim, H.W.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
7. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
8. Park, N.: Secure UHF/HF Dual-Band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
9. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. International Journal of Ad Hoc and Ubiquitous Computing 8(4), 205–219 (2011)
10. Park, N., Kim, Y.: Harmful Adult Multimedia Contents Filtering Method in Mobile RFID Service Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS (LNAI), vol. 6422, pp. 193–202. Springer, Heidelberg (2010)
11. Park, N., Song, Y.: AONT Encryption Based Application Data Management in Mobile RFID Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS (LNAI), vol. 6422, pp. 142–152. Springer, Heidelberg (2010)
12. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
13. Park, N.: Secure Data Access Control Scheme Using Type-Based Re-encryption in Cloud Environment. In: Katarzyniak, R., Chiu, T.-F., Hong, C.-F., Nguyen, N.T. (eds.) Semantic Methods. SCI, vol. 381, pp. 319–327. Springer, Heidelberg (2011)
14. Park, N., Song, Y.: Secure RFID Application Data Management Using All-Or-Nothing Transform Encryption. In: Pandurangan, G., Anil Kumar, V.S., Ming, G., Liu, Y., Li, Y. (eds.) WASA 2010. LNCS, vol. 6221, pp. 245–252. Springer, Heidelberg (2010)
15. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. The Journal of Korea Information and Communications Society 35(5), 785–793 (2010)
16. Kim, Y., Park, N.: Development and Application of STEAM Teaching Model Based on the Rube Goldberg's Invention. In: Yeo, S.-S., Pan, Y., Lee, Y.S., Chang, H.B. (eds.) Computer Science and its Applications. LNEE, vol. 203, pp. 693–698. Springer, Heidelberg (2012)
17. Park, N., Cho, S., Kim, B., Lee, B., Won, D.: Security Enhancement of User Authentication Scheme Using IVEF in Vessel Traffic Service System. In: Yeo, S.-S., Pan, Y., Lee, Y.S., Chang, H.B. (eds.) Computer Science and its Applications. LNEE, vol. 203, pp. 699–705. Springer, Heidelberg (2012)

Author Index